UNIVERSITE DE PARIS X-NANTERRE

U.F.R. DE PHILOSOPHIE

JEAN LASSEGUE

L'intelligence artificielle
et
la question du continu

Remarques sur le modèle de Turing

THESE DE PHILOSOPHIE

en vue du

DOCTORAT DE L'UNIVERSITE DE PARIS X-NANTERRE

préparée sous la direction de

Monsieur le professeur Daniel ANDLER

«Ne soyons pas si puristes et soyons reconnaissants au continu qui, si *tout* sort du nombre entier, était seul capable d'en faire *tant* sortir».

Henri Poincaré, *La valeur de la science*, p. 150.

Remerciements

En repensant au chemin parcouru pendant les années qui ont vu mûrir mon projet de thèse, c'est tout d'abord à mon frère, Bernard Lassègue, que vont mes remerciements. C'est lui qui, en me montrant ses premiers programmes écrits sur cette antiquité qu'est aujourd'hui l'Apple IIc, m'a fait sentir que c'est moins l'autorité - transitoire - d'un paradigme scientifique qui devait retenir mon attention d'épistémologue que la satisfaction spécifique que procure l'attitude scientifique en elle-même. Il a par la suite soutenu matériellement mon travail en m'envoyant des États-Unis les nombreux livres et articles que j'aurais eu le plus grand mal à trouver en France.

J'ai pu observer la satisfaction que procure l'attitude scientifique en d'autres occasions. Je pense à David Freedman, à nos après-midis d'algèbre à Oxford et à nos conversations sur Leibniz. Je pense aussi à Driss Abouabdillah qui a eu la patience, à Rabat, d'exercer avec moi son art de mathématicien : je me rappelle avec plaisir nos soirées où, abandonnant les exercices sur les pavages achevés, décidément trop difficiles pour moi, il lisait puis me traduisait Ibn Arabi. Je pense enfin tout particulièrement au dialogue avec Jacques Jeanjean dont les talents mathématiques et philosophiques conjugués m'ont apporté, durant toutes ces années, l'émulation nécessaire à la poursuite de mon travail.

Une place toute particulière doit être réservée aux remerciements que je dois à Jean-Michel Salanskis, l'une des rares personnes pour qui la transmission du savoir soit une vertu naturelle. C'est lui qui, voici quatre ans, a donné forme à mon projet de thèse et qui, depuis lors, n'a cessé de relire mon travail en respectant mes choix théoriques, qu'il ne partage pas toujours. Je lui dois beaucoup, tant pour le savoir que j'ai acquis auprès de lui que pour la liberté qu'il manifeste dans sa façon de le dispenser.

J'ai mis un certain nombre de mes amis à contribution et je les remercie d'avoir répondu à mes demandes avec bonne grâce : Pierre Gervais qui le premier a corrigé un certain nombre d'erreurs touchant mon interprétation du jeu de l'imitation; Andrew Gumbel qui m'a donné force détails touchant la civilisation britannique; Nicolas Michel pour la finesse de ses critiques; Alexis Tadié à qui est revenue la tâche ingrate de la lecture du manuscrit et dont les conseils à la fois sur le fond et sur la forme m'ont beaucoup servi. Je remercie aussi mon maître en philosophie, Serge Boucheron, pour l'attention qu'il a portée à la lecture de mon travail et l'acuité de ses critiques. Les conversations que j'ai pu avoir avec Yves-Marie Visetti enfin m'ont été d'un précieux secours.

J'ai bénéficié au cours de ces années de l'aide de plusieurs institutions. Les deux ans passés au CNRS au sein du laboratoire du Crea m'ont apporté une aide matérielle et intellectuelle décisive : je remercie tout particulièrement Daniel Andler qui fit le pari que mon sujet de thèse pouvait avoir un intérêt; Jean-Pierre Dupuy, Jean Petitot, André Orléans et Bénédicte Reynaud pour leur soutien. Ma dernière année, passée à la Fondation Thiers, m'a permis d'achever mon travail dans de bonnes conditions.

Contrairement à l'habitude qui consiste à reléguer dans l'anonymat de la biographie de fin de volume tous les auteurs pêle-mêle, je voudrais mentionner quatre livres qui ont nourri ma réflexion : il s'agit du *Recherches sur une logique de la pensée créatrice en mathématiques* de Maurice Meigne; du *From Mathematics to Philosophy* de Wang Hao; du *Principes classiques d'interprétation de la nature* de Jean Largeault et du *Alan Turing, The Enigma of Intelligence* de Andrew Hodges. C'est aux auteurs de ces livres que j'aimerais témoigner de ma reconnaissance.

C'est enfin à ma famille que vont mes remerciements : à mes parents tout d'abord qui m'ont soutenu matériellement et moralement tout au long de ces années; à Mouna ensuite qui m'a initié aux mystères de l'intelligence des enfants; à Tourya enfin sur qui je me repose tant pour la bonne marche de ma vie. C'est à elle que je dédie ce travail.

Avant-propos

L'expression "d'intelligence artificielle" peut être interprétée de façon plus ou moins large. Au sens étroit, il s'agit d'un projet technologique qui rend possible une connaissance positive de l'esprit humain. Au sens large - que nous adoptons ici - l'expression désigne plutôt une conception philosophique qui émane du formalisme logique de la machine de Turing et qui utilise celui-ci comme un outil conceptuel pour penser la notion d'esprit. L'idée philosophique d'intelligence artificielle reprend ainsi sous une forme nouvelle un problème classique, celui de la possibilité de concevoir objectivement une "science de l'esprit". Quelle est la nature de la difficulté rencontrée ?

La difficulté provient de ce que, pour concevoir une science de l'esprit, il faudrait réussir à envisager le phénomène de la pensée comme de l'extérieur. Une telle attitude présuppose qu'il est possible, tout en étant immergé dans la pensée, d'en délimiter ce qu'il faudrait appeler les "bords extérieurs". Il y a là une difficulté qui provient de ce qu'on ne voit pas par quel moyen on pourrait délimiter de l'extérieur la pensée, parce qu'il semble impossible de s'en détacher[1] : c'est cette adhérence de la pensée à elle-même que nous caractérisons par le biais de la notion de *continuité* et qui apparaît comme un obstacle majeur pour qui veut décrire sur un mode objectif la nature de la pensée.

[1] Comme le fait remarquer R. Ruyer : «L'"absence de bords" vaut pour l'ensemble d'une conscience individuelle. Nous ne nous sommes pas vus naître et nous ne nous verrons pas mourir. Nous ne nous voyons pas nous endormir, ou nous évanouir. Nous ne sommes pas conscients d'oublier. La conscience est vision d'elle-même, mais elle ne se voit pas comme partie d'un tout (sauf par des procédés symboliques toujours imparfaits et indirects). Ce paradoxe est aussi important pour la situation humaine en général que pour le statut de la connaissance humaine. Notre vie consciente n'est bordée par rien.» **R. Ruyer**, *Paradoxes de la conscience et limites de l'automatisme*, Albin Michel, Paris, 1966, p. 17.

1. Éclaircissement du sujet

L'intelligence artificielle permet-elle d'apporter une réponse à ce problème en proposant un modèle de l'esprit qui se veut objectif ? Elle délimite du moins un domaine d'objectivité et une méthode d'investigation qui lui sont propres.

On peut dire que, dans le cas de l'intelligence artificielle, le domaine d'objectivité ainsi que la méthode d'investigation sont dominés par une perspective essentiellement *discrète* : on considère en effet que l'esprit possède des états qui sont isolables les uns des autres et qui sont susceptibles d'être représentés par des signes dont les relations mutuelles sont régies par les mêmes règles que celles que l'on utilise pour manipuler les entités d'un système formel. Étudier ces signes et leurs rapports pour eux-mêmes, en tant que ces derniers sont susceptibles d'être porteurs d'une objectivité, voilà ce qui doit constituer, s'agissant de l'intelligence artificielle, le domaine propre d'une science de l'esprit[2] : le rôle des signes et de leur traitement formel serait de permettre de franchir l'obstacle de l'adhérence de la pensée à elle-même, ce que nous avons appelé à l'instant sa "continuité". Aussi cette science dépend-elle d'un niveau d'objectivité spécifique appelé "niveau informationnel", celui du *traitement calculatoire des signes*.

Outre le fait qu'un tel niveau n'a jamais été isolé expérimentalement et que son existence reste dès lors l'objet d'une pure spéculation, l'expression de "niveau informationnel" regroupe plusieurs hypothèses, que l'on pourrait décomposer en trois sous-ensembles[3] : penser (dans un sens suffisamment large pour inclure la perception) implique de traiter de *l'information*; le traitement de

---

[2] C'est Helmholtz qui, le premier, semble avoir pensé le rapport à la nature sur le mode d'un traitement symbolique : «Les yeux et le nerf optique sont aveugles. Les sensations visuelles doivent être interprétées comme des signes du monde extérieur. Cette interprétation est le résultat du caractère a priori des fonctions intellectuelles» cité dans **T. C. Meyering**, *Historical roots of cognitive science*, *the rise of a cognitive theory of perception from Antiquity to the 19th century*, Kluner Academic Publishers, Dordrecht, 1989, p. 216.

[3] Cf. **D. Andler,** "Les sciences de la cognition" dans [**J. Hamburger dir.**, *La philosophie de sciences aujourd'hui*, Gauthier-Villars, Paris, 1986], p. 135.

l'information proprement dit est le résultat d'un *calcul* sur des signes; ces signes sont eux-mêmes porteurs de *sens*. L'hypothèse de l'existence d'un traitement de l'information sous forme d'un calcul permet de *simuler* sur ordinateur ce que l'on suppose correspondre au niveau informationnel servant de soubassement à la vie psychologique de la pensée. Ceci définit le projet et la démarche adoptés par l'approche symbolique et calculatoire en intelligence artificielle - le "cognitivisme" -, seule approche qui nous retiendra ici. Pourquoi vouloir, dans ces conditions, introduire une réflexion sur le continu en intelligence artificielle ?

Remarquons tout d'abord que le *continu* n'est pas la *continuité*. Comme nous l'avons vu, la continuité est une propriété subjective qui permet de caractériser l'adhérence de la pensée à elle-même. En revanche, la notion de continu permet de décrire des entités en les délimitant; c'est le cas pour un mouvement considéré comme continu ou pour un objet comme un segment d'une droite géométrique [4]. En ce sens, le continu possède bien un contenu objectif susceptible d'une détermination décrite dans les signes du langage. Cependant, le continu entretient avec la notion de continuité des liens qui font de lui un objet mystérieux, dans la mesure où il apparaît comme *en excès* par rapport à toute détermination langagière, de nature essentiellement discrète. C'est cet excès qui apparaît comme une trace de la parenté du continu avec la notion de continuité, qui nous était apparue comme "sans bords". Aussi le continu, tout en étant une entité déterminable objectivement et recevant, de ce fait, un statut mathématique, entretient avec la pensée telle qu'elle se donne subjectivement dans son adhérence à elle-même, des liens secrets.

On comprend dès lors comment la confrontation entre le cadre discret de l'intelligence artificielle et la notion de continu peut commencer à faire sens : grâce à elle, il devient possible de s'interroger sur la légitimité du modèle discret de la pensée proposé par l'intelligence artificielle. En effet, la notion de continu, de par son rapport à la continuité, permet de reposer le problème de l'adhérence de

---

[4] Nous reprenons à M. Panza la distinction de la continuité comme propriété et du continu comme objet. Cf. **M. Panza**, "De la continuité comme concept au continu comme objet", dans [*Le Labyrinthe du Continu*, **J.-M. Salanskis** et **H. Sinaceur dirs.**, Springer-Verlag, Paris, 1992], p. 18.

la pensée à elle-même et de voir si la solution du “modèle discret” telle qu’elle est proposée par l’intelligence artificielle pour en rendre compte est ou non une solution satisfaisante.

2. Point de vue philosophique adopté

Dans le cadre de l’intelligence artificielle, le continu n’apparaît habituellement que comme le contraire du discret. Aussi, dans une perspective dominée par le discret comme c’est le cas en intelligence artificielle, tout ce qui apparaît comme continu, objet ou processus, est-il susceptible d’être *réduit* à du discret.

21. Critique du réductionnisme dualiste

Dans cette optique, le continu n’apparaît au mieux que comme un terme vide toujours “discrétisable” et qui n’est, de ce fait, porteur d’aucune objectivité. Cette réduction toujours possible du continu au discret est liée à la méthodologie de l’intelligence artificielle et à sa façon de concevoir son domaine d’objectivité, c’est-à-dire son rapport à la réalité : si l’on peut réduire le continu au discret, c’est que la méthodologie de l’intelligence artificielle est conçue comme une forme abstraite seule porteuse d’objectivité, forme qui vient s’appliquer à un divers qu’il s’agit, au sens propre, d’informer. Forme et matière une fois séparés, leur articulation devient difficile, voire impossible, à penser : aussi des concepts, et au premier chef celui de continu, dépossédés de tout statut légal par la méthodologie discrète de l’intelligence artificielle, apparaissent-ils réduits à n’être que des termes métaphysiques informes et perdent-ils tout fondement objectif.

Mais il est possible d’envisager autrement le rapport de l’intelligence artificielle à la réalité, dans une perspective qui ne soit pas dualiste. Le dualisme de l’intelligence artificielle provient de son caractère instrumentaliste : les concepts sont conçus, dans cette interprétation, comme des outils qu’il faut appliquer à la réalité pour la constituer objectivement. Le paradigme de la machine - outil par excellence - qui domine ce champ du savoir est très largement responsable de cette conception. Cependant, si l'on part de l'idée que les concepts ne sont pas ce *au moyen de quoi* on pense la réalité, mais *ce qui en est intelligible*, alors le continu est, au même titre que le discret, une émanation de cette réalité et peut reprendre un sens objectif qu'il semblait avoir définitivement perdu.

22. La notion de représentation et sa théorie objective

Le fondement d'une science de l'esprit tel qu'il est envisagé par le cognitivisme repose sur une théorie de la représentation basée sur la méthodologie discrète de l'intelligence artificielle. Le terme de "représentation", chargé d'une longue histoire philosophique - avant d'avoir un sens mathématique -, désigne avant tout un rapport : la représentation apparaît en effet comme un couple constitué du représentant (sujet connaissant) et du représenté (objet), les deux termes n'ayant pas d'existence hors de la représentation. Une théorie de la représentation a pour mission de décrire comment s'établit le rapport du sujet connaissant à l'objet. Dans le cas de l'intelligence artificielle, ce rapport est décrit comme une simulation des objets au moyen de leurs représentés. Ces représentés sont, on l'a vu, des signes discrets dont les rapports font l'objet d'un calcul. La simulation consiste donc à opérer un calcul sur des signes. Deux difficultés apparaissent alors pour une théorie de la représentation qui se voudrait objective.
Premièrement, l'utilisation du calcul a pour but d'éliminer d'une telle théorie tout l'aspect subjectif qui est lié à la notion de représentation, c'est-à-dire la façon dont la représentation laisse entrevoir l'existence de facultés psychologiques chez le sujet connaissant. Pour qu'une telle démarche soit légitime, il faut considérer que la notion de calcul, bien qu'elle soit manipulée par des êtres humains, est entièrement à l'abri des investissements subjectifs habituellement nécessaires chez eux à la constitution d'un *intérêt* psychologique, parce que sinon la théorie n'aurait plus de fondement objectif. Il y a là un présupposé touchant la notion de calcul dont il faut mesurer la portée, car on sent immédiatement la difficulté qu'il y a à conférer à une notion le rôle de fondement sans que celle-ci soit investie par l'intérêt psychologique de ceux qui l'utilisent pour cette fin. Or la notion d'intérêt

psychologique n'a évidemment rien d'objectif, puisqu'elle définirait plutôt en propre la sphère du subjectif[5].

Deuxièmement, en faisant l'hypothèse que l'objectivité de la théorie de la représentation dépend du calcul sur des représentés, on tend à occulter la façon dont à des représentés correspond une réalité qui leur est étrangère et qui leur sert de fondement. D'où vient qu'on soit certain que la simple simulation des représentés suffise à assurer un lien avec un "quelque chose" non-représenté émanant de la réalité ? On objectera que cette question est inutile puisqu'on a pris garde de dire que le représenté n'a pas d'existence hors du *rapport* de représentation. Mais si c'est bien quand il entre dans un rapport que le représenté devient objet de connaissance, on fera remarquer que le premier rapport au sein duquel il surgit est à l'évidence le rapport à la *réalité* dont il émane. En considérant la notion de calcul comme un fondement objectif, le rapport du représenté à ce qu'il représente devient obscur pour une raison semblable à celle qui avait occulté la référence à la psychologie du sujet connaissant. Ces deux occultations sont en effet liées : il ne peut y avoir de représenté que par suite d'un déplacement psychologique qui favorise l'intériorisation de l'objet extérieur en une entité signifiante pour le sujet et qui constitue du même coup l'objet en signe. Ce déplacement psychologique exige donc que le sujet y soit *intéressé*. C'est ce processus psychologique qui doit pouvoir être pris en compte par toute théorie de la représentation qui voudrait se constituer en théorie objective.

Une théorie calculatoire de la représentation comme l'intelligence artificielle a donc de la difficulté à penser conjointement le rapport du représentant à ce qu'il représente d'une part et le rapport du représenté à ce qu'il représente d'autre part, parce qu'elle met principalement l'accent sur le *rapport* de représentation décrit en termes calculatoires. Il y aurait ainsi une élimination

---

[5] Le mathématicien Alan Turing le faisait, à sa façon, remarquer quand il tentait de décrire les facultés mises en jeu dans le raisonnement mathématique : «On peut caractériser grossièrement le raisonnement mathématique par l'exercice conjoint de deux facultés, que nous pouvons appeler l'*intuition* et l'*ingéniosité*. [note : Nous ne prenons pas en compte une faculté des plus importantes, celle qui distingue les sujets d'importance des autres]». **A. M. Turing**, "Systems of Logic based on Ordinals", Proceedings of the London Mathematical Society, ser. 2, vol. 45 (1939), pp. 161-228, réédité dans [**M. Davis ed.**, *The Undecidable*, Raven Press, Hewlett, New York, 1965], pp. 154-222.

conjointe du sujet connaissant et du fondement réel du représenté qui proviendrait du même présupposé : accorder un caractère *abstrait* à la notion de calcul. C'est cette abstraction qui tend en effet à occulter le problème de savoir comment un sujet peut s'investir dans la représentation et comment un objet de la réalité peut devenir un représenté qui fasse office de signe pour un sujet. Bref, en visant à remplacer la notion de faculté subjective par celle de calcul et en limitant la théorie de la représentation à un traitement des représentés, une théorie objective de la représentation comme l'intelligence artificielle tend à laisser dans l'ombre la façon dont les sujets constituent psychologiquement des objets en représentés.

Nous faisons l'hypothèse que la notion de continu permet d'envisager autrement les deux difficultés dont nous parlions à l'instant *en remettant en question le caractère abstrait attribué à la notion de calcul*.

23. Le calcul et le continu

Le continu a ceci de particulier qu'il n'est pas le contraire du discret mais son *autre*. Aussi, dans une théorie discrète et calculatoire de la représentation comme l'intelligence artificielle, aucun signe ou ensemble de signes faisant office de représenté discret ne peut-il référer adéquatement au continu : la notion de continu excède son représenté parce qu'elle excède les pouvoirs du calcul. Ce faisant, elle excède ce qui sert de fondement à une théorie objective et calculatoire de la représentation comme l'intelligence artificielle. Cet excès est la manifestation d'une réalité qui ne tire pas son origine du rapport interne à la représentation.

Dès lors, la notion de continu permet d'éclairer les deux rapports laissés dans l'ombre par une théorie objective de la représentation et qui lui sont pourtant nécessaires, le rapport du représentant au sujet connaissant d'une part et le rapport du représenté à la réalité de l'autre. Aussi cherchera-t-on à soutenir que ce par quoi une science discrète des représentations comme l'intelligence artificielle est *objective* provient non pas du caractère abstrait de la notion de calcul mais de ce qu'elle doit en même temps nier dans sa méthode et présupposer dans son ontologie : qu'il y ait un fondement continu à la représentation. Nous tenterons donc de montrer que ce qui permet de constituer l'intelligence artificielle en une théorie objective de la représentation implique d'y intégrer la notion de continu parce que c'est elle qui en nourrit l'objectivité.

C'est d'ailleurs ce dont on aurait pu s'apercevoir directement en usant d'un point de vue historique : *l'acte de naissance du projet d'intelligence artificielle fut en effet une réflexion sur les rapports du continu et du discret.* C'est à propos de la question de la calculabilité des nombres réels que le mathématicien britannique Alan Turing finit en effet par concevoir, entre la fin des années trente et le début des années cinquante de ce siècle, la possibilité d'une "intelligence artificielle"[6]. Aussi est-ce l'œuvre de Turing et tout particulièrement son invention d'un concept original de machine qui nous servira, tout au long de cette étude, de fil directeur.

24. Philosophie de la nature

Le point de vue philosophique adopté dans cette étude n'est donc pas de type instrumentaliste mais bien plutôt celui d'une philosophie de la nature[7]. Dans cette optique, le discret et le continu ne sont pas deux termes qui s'opposent radicalement mais ils entretiennent des rapports qui font partie intégrante de la notion de signification. C'est ce que faisait remarquer R. Thom[8] : une structure mathématique discrète doit être plongée dans un continu pour avoir,

---

[6] L'expression d'intelligence artificielle n'est cependant pas de Turing. L'expression a pour origine un programme de recherche élaboré en 1956 par un groupe d'informaticiens lors d'un séminaire à Dartmouth, aux États-Unis et elle semble avoir été forgée à cette occasion par John McCarthy. L'expression s'est popularisée grâce à un article de Marvin Minsky écrit et divulgué en 1957 mais publié seulement en 1963, appelé "Steps toward Artificial Intelligence" publié dans [*Computers and Thought*, **Feigenbaum** et **Feldman eds.**, New-York, 1963]. Cf. **H. Gardner,** *The Mind's New Science*, Basic Books, New-York, 1985, p. 30 et 139.

[7] La philosophie de la nature apparaît, au vingtième siècle, comme à peu près inexistante dans le monde anglo-saxon, exception faite, en Angleterre, de D'Arcy Thompson et de A. N. Whitehead; il s'agit d'une tradition qui reste largement minoritaire dans ce qu'il est convenu d'appeler la "philosophie continentale", mais qui est bien représentée aujourd'hui en France, grâce aux travaux de R. Ruyer, G. Simondon, R. Thom et J. Largeault. Cf., pour des remarques historiques touchant cette tradition, **J. Largeault**, *Principes de philosophie réaliste,* Klincksiek, Paris, 1985, pp. 263-264 et *Principes classiques d'interprétation de la nature*, Publications de l'Institut Interdisciplinaire d'Etudes Epistémologiques, Paris, Vrin, 1988, p. 429.

[8] Cf., par exemple, **R. Thom**, *Paraboles et catastrophes*, Flammarion, Paris, 1983, pp. 146-147. R. Thom prend l'exemple de la notion d'infini non-dénombrable qui ne revêt pour nous une signification que si on la rapporte mentalement à un segment de droite géométrique.

psychologiquement, un sens. La constitution d'une intelligibilité passe par cette opération mentale qui consiste à plonger les concepts discrets dans le continu : c'est par ce biais que le concept discret acquiert une délimitation stable. La genèse des concepts et de leur pouvoir descriptif implique ainsi de faire appel au continu comme à ce qui rend leur intelligibilité possible. Le continu, de ce fait, a rang de *principe d'intelligibilité.* Plus précisément, c'est notre capacité d'expression de l'intelligibilité des phénomènes qui rend nécessaire un appel à la notion de continu. Tâchons de comprendre pourquoi.

La possibilité d'une image stable et linguistiquement exprimable des phénomènes internes (intuition du temps) comme des phénomènes externes (appréhension d'une réalité) nécessite de concevoir les objets à la fois comme essentiellement distincts du langage et comme susceptibles cependant d'être linguistiquement décrits. Or cette capacité à forger des images stables et linguistiquement descriptibles n'est pas arbitraire et ne ressort pas de nos simples capacités subjectives de penser : elle est un effet provoqué sur notre organisme par la nature à l'extérieur de nous, *nature dont nous faisons cependant partie*. C'est là tout le paradoxe de la réflexivité[9] : c'est notre position d'organismes naturels, à la fois dans la nature mais possédant la capacité, comme tout organisme autonome, d'opérer une distinction entre la nature (pensée comme extérieure) et nous-mêmes (pensés comme ayant un intérieur) qui nous rend capables de nous *représenter* sous la forme de contenus de pensée la nature et nous-mêmes. Les phénomènes possèdent la même cohérence que celle de nos organismes : ce qui nous apparaît comme ayant une délimitation stable dans l'espace-temps se trouve, comme nous-mêmes, à la fois distingué et plongé dans la nature. Toute représentation des phénomènes, internes et externes, relève de ce mouvement qui consiste à délimiter une partie de la nature tout en y étant plongé : aussi est-ce la notion même de représentation qui exige d'user de la notion de continu puisque celle-ci est intuitivement décrite comme ce dans quoi nous

---

[9] Cf. **R. Ruyer**, *Paradoxes de la conscience et limites de l'automatisme*, Albin Michel, Paris, 1966, pp. 68 : «Du fait que, dans notre champ visuel, nous entrevoyons l'"avant" de notre propre corps, nous imaginons que nous voyons notre corps et même presque nos yeux (réduits à nos cils, devinés et complétés). Nous voyons (presque) nos yeux voir les autres objets».

sommes plongé mais dont nous embrassons cependant des parties. C'est donc bien la capacité à posséder des représentations qui implique de supposer la présence d'un continu comme cause des effets du monde extérieur sur nos organismes. La possibilité d'une description de tout phénomène interne et externe consiste donc à reconnaître que le phénomène n'apparaît que sous la modalité de la distinction entre l'intérieur et l'extérieur, c'est-à-dire de ce qui est plongé dans une totalité tout en s'en distinguant. *La notion de continu apparaît ainsi comme inhérente au schème corporel dont sont issues les représentations.* L'aspect linguistique de la description des phénomènes présuppose donc la constitution d'une représentation qui, en son fond, repose sur la notion de continu. C'est donc finalement la distinction entre l'intérieur et l'extérieur qui permet l'accès à une description linguistiquement stable de la nature et de nous-mêmes en tant que partie de la nature.

Aussi quand on étudie la notion de continu, n'est-ce pas la réalité, pensée comme essentiellement étrangère à nous-mêmes et comme vide de toute détermination que l'on suppose continue mais les processus causaux qui nous la rendent intelligible. C'est cet aspect causal et "réflexif" du continu, en tant qu'il est principe d'intelligibilité des formes dans la nature qui nous retiendra au cours de cette étude, parce que c'est par ce biais qu'une analyse de nos processus de pensée devient possible. C'est en particulier sur les liens secrets qui existent entre la notion de continu et notre capacité d'invention qu'il faudra s'interroger[10].

On comprend mieux dès lors le titre de notre travail : si l'intelligence artificielle est bien liée à une méthodologie discrète, alors que peut-elle dire de la constitution des processus de pensée qui impliquent l'usage de la notion de continu? On empruntera deux directions pour tenter de répondre à cette question. Celle de la constitution des représentations tout d'abord; on tentera alors de répondre à la question de savoir comment une méthodologie discrète peut rendre compte de cette constitution. Celle de la constitution du modèle discret de

[10] Cette capacité d'invention liée à la capacité à se représenter les phénomènes apparaît aussi dans le domaine mathématique lui-même. Le "problème du continu", au sens technique, avait été constitué dès l'avènement de la théorie cantorienne des ensembles. Il énonce sous forme de conjecture qu'il n'y a pas de cardinalité intermédiaire entre l'infini dénombrable et la puissance du continu.

l'activité de l'esprit ensuite : si la constitution du modèle discret en question implique l'usage de la notion de continu, comment ce modèle discret peut-il rendre compte de sa propre émergence, c'est-à-dire quelle peut être sa propre capacité à opérer un réflexion sur soi, une "pensée" ?

Le point de vue de la philosophie de la nature permet d'aborder ces questions sous un angle d'attaque nouveau. Il faut, pour ce faire, revenir à la méthodologie de l'intelligence artificielle et envisager autrement la notion de calcul. En s'appuyant sur une réflexion du mathématicien Emil Post[11], on peut considérer que la notion de calcul n'est pas une définition ou un axiome abstraits, c'est-à-dire une notion formelle relevant de la logique et qu'il faudrait appliquer à la "réalité" mais bien plutôt une *loi de la nature*. Selon notre point de vue, le concept de calcul ne peut dès lors avoir la signification d'une loi de la nature que s'il est rapporté au continu. Dans ce cas, l'intelligence artificielle qui est une manifestation de cette loi, doit elle aussi y être rapportée, si l'on veut rendre compte de sa genèse. On comprend mieux dès lors pourquoi l'acte de naissance de l'intelligence artificielle est, selon nous, lié à une réflexion sur les rapports du continu et du discret : c'est en effet par ce biais qu'elle a pu devenir une science, comme nous tenterons de le montrer au cours de ce travail.

3. Exposition du projet

On comprend dès lors en quel sens il faut entendre le titre de ce travail : la notion de continu joue le rôle d'une *archéologie de la représentation* qui doit mener, par-delà la question du rapport calculatoire du représentant au représenté, à la question ontologique du fondement objectif du représenté. Examinons la façon dont nous allons mener cette enquête.

31. La psychologie dans la logique

[11] Cf. **E. Post** (1936), "Finite Combinatory Processes - Formulation I", Journal of Symbolic Logic, I, réédité dans [**M. Davis ed.**, *The Undecidable*, Raven Press, Hewlett, New York, 1965], pp. 288-291.

La première partie de notre travail, intitulée “La psychologie dans la logique”, vise à comparer deux présentations de la notion de calculabilité quand on use du formalisme de la machine de Turing.

Tel qu’on le présente habituellement, c’est l’équivalence du formalisme de la machine de Turing avec d’autres formalismes classiques portant sur le même objet, le calcul, qui corrobore l’objectivité de ce formalisme. De façon plus spécifique, c’est en se reposant sur l’équivalence formelle avec d’autres formalismes que le formalisme de la machine de Turing est utilisé pour servir de fondement à la théorie de la représentation en intelligence artificielle. Je tâcherai, pour exposer ce point de vue, de présenter le formalisme de la machine de Turing comme on le présente habituellement et de montrer ensuite en quel sens il peut être considéré comme formellement équivalent au formalisme de la théorie des fonctions récursives.

Je montrerai ensuite que la profonde originalité de la présentation par Turing du formalisme que l’on appellera par la suite la “machine de Turing” vient non pas tant de son objectivité reposant sur son équivalence formelle avec d’autres formalismes que de ce *qu’il introduit en logique des considérations psychologiques*: c’est en effet en faisant appel à la notion idéale d’un “calculateur” (dénommé en anglais *computer*) que Turing présente son concept de machine et expose par ce biais ce qu’il faut entendre par calculabilité.

32. La logique dans la psychologie

Ma deuxième partie qui s’intitule “La logique dans la psychologie” étudie le déplacement du concept de machine de Turing vers la psychologie, c’est-à-dire le déplacement inverse de celui de la première partie.

On justifie habituellement le déplacement du formalisme logique de la machine de Turing vers la psychologie en faisant appel à l’objectivité et à l’universalité du formalisme en question. La capacité universelle de ce formalisme à *simuler* tout phénomène suffirait à justifier son application au cas des états mentaux. Il n’y aurait pas de spécificité propre au cas de la psychologie, que l’on pourrait traiter comme tout autre objet de la nature.

J'examinerai s'il est légitime de s'appuyer sur l'objectivité logique de la notion de calcul et sur l'universalité du formalisme de la machine de Turing pour s'autoriser à déplacer le concept de machine de Turing hors de son domaine originel qui est celui de la théorie du calcul, vers celui de la psychologie. Car si l'introduction de données psychologiques en logique peut trouver une justification, l'inverse pose un problème philosophique majeur : Turing était bien conscient du problème, lui qui introduisit, comme nous venons de le voir, des considérations psychologiques dans la présentation logique du formalisme de la machine de Turing[12] sans qu'une telle démarche lui paraisse problématique, mais qui a consacré tout un article à la question de la possibilité d'une introduction de ce formalisme logique en psychologie[13]. En effet, deux conceptions philosophiques divergentes s'en suivent, l'une instrumentale et l'autre naturelle. Je montrerai que les deux conceptions philosophiques se superposent dans l'article de Turing qui traite de la question des rapports de la logique à la psychologie, article que l'on a tendance habituellement à n'interpréter que d'un point de vue instrumental.

Si l'on s'appuie sur l'objectivité de la notion de calcul pour s'autoriser à la déplacer dans le contexte de la psychologie, deux conséquences majeures s'en suivent.

Premièrement, on est amené à concevoir l'objet de cette psychologie comme une matière susceptible de recevoir une forme, celle d'un traitement de nature calculatoire. On fait donc une hypothèse instrumentaliste sur la nature de l'esprit en supposant qu'il peut recevoir un traitement de ce type. Or une attitude instrumentaliste a pour conséquence une attitude dualiste en philosophie, dualisme qui tend à laisser sans réponse les questions liées au fondement de l'objectivité et à la vraie nature de la représentation.

Deuxièmement on tend à considérer le concept de machine de Turing comme un fondement pour la psychologie. La conséquence la plus directe est que

---

[12] **A. M. Turing** (1936), 'On Computable Numbers with an Application to the Entscheidungproblem', Proceedings of the London Mathematical Society, 42, pp. 230-265.
[13] **A. M. Turing** (1950), "Computing Machinery and Intelligence", Mind, vol LIX, n°236, pp. 433-460.

l'on occulte la question de l'origine du concept, c'est-à-dire la façon dont il a été lui-même psychologiquement *inventé*. Le formalisme de la machine de Turing possède lui aussi, comme tout concept, une origine psychologique et le fait qu'on puisse l'appliquer à la psychologie ne règle pas la question de son origine psychologique. Il faut alors répondre à la question suivante, qui se présente comme un cercle vicieux : quelle est l'origine psychologique du formalisme qui sert à fonder la psychologie ? Ou encore : par quel moyen étudier la proto-psychologie du concept qui sert de fondement à la psychologie ? C'est en tentant de répondre à ces questions qu'il devient possible de prendre en considération la notion de continu dans son sens substantif et non pas seulement dans son sens "adjectival"[14].

33. La portée générale du modèle de Turing

J'examinerai alors dans une troisième partie comment il est possible d'exploiter cette interprétation non-instrumentale de l'application du formalisme de la machine de Turing à la psychologie pour constituer une théorie non-dualiste de la représentation. Cette théorie devra s'appuyer sur des considérations touchant à la notion de continu puisque c'est cette notion qui nous est apparue comme le fil directeur d'une enquête sur l'objectivité de la notion de représentation. Nous tenterons d'examiner s'il est possible de mettre au jour des contraintes objectives qui informent les représentations. En particulier, nous tenterons de montrer comment la différence entre l'intérieur et l'extérieur du corps, biologiquement fondée, a un rôle contraignant sur les formes des représentations. Il sera alors possible d'étudier les conséquences de cette théorie d'un point de vue linguistique et logique.

4. Objections et réponses aux objections

[14] Nous nous inspirerons des remarques de G. Simondon contenues dans le chapitre "L'invention technique : fond et forme chez le vivant et dans la pensée inventive" qui nous paraît tout à fait fondamental du point de vue d'une philosophie de la nature. Cf. **G. Simondon**, *Du mode d'existence des objets techniques,* Aubier, Paris, 1958, pp. 56-60.

Certaines objections de principe peuvent être formulées concernant le projet même de ce travail et nous voudrions les dissiper avant d'entamer notre réflexion proprement dite. Elles sont de deux ordres, l'une scientifique, l'autre phénoménologique.

41. Objections scientifiques

On peut tout d'abord se demander ce que le continu tel qu'il est décrit en mathématique et tel qu'il est utilisé en physique a à voir avec le continu tel que je l'ai décrit dans le cadre d'une recherche épistémologique sur les conditions d'objectivité d'une science de l'esprit : quel rapport entre le continu logico-mathématique d'une part et la continuité psychologique d'autre part ? On répondra en faisant remarquer que cette question revient à se demander s'il est légitime de rapprocher la logique de la psychologie. Mais c'est le *projet même de l'intelligence artificielle qui opère ce rapprochement* du point de vue de la notion discrète de calcul. Dès lors, il n'est pas interdit d'envisager le même rapprochement du point de vue du continu, si l'on estime, comme je le fais, qu'une science de l'esprit, pour se constituer en science, doit nécessairement en passer par là.

Des objections plus précises peuvent être formulées du point de vue du continu mathématique et du point de vue du continu physique, une fois que l'objection de principe touchant au rapprochement des notions de continu et de continuité a été levée.

411. Objection mathématique

On peut objecter qu'il n'y a pas de raison d'accorder à la notion de continu un statut si particulier, alors qu'elle n'est qu'une notion parmi d'autres dans le domaine logique et mathématique. Examinons un instant cette notion.

Au cours de ce siècle, la logique mathématique s'est principalement heurtée à la question du continu sous l'aspect de la question ensembliste de sa *cardinalité* et ce, depuis l'avènement de la théorie cantorienne des ensembles[15]. La question de la cardinalité du continu s'exprime d'une manière fort simple : combien y-a-t-il de points dans un intervalle donné de la droite réelle[16] ? Cette question institue, dans le cas du continu, un régime de représentation à la fois géométrique (il est question de la notion de droite) et numérique (il est question des nombres réels qui "composent" la droite géométrique quand ils sont associés à ses points[17]). En tant que telle, cette représentation, au fondement de la géométrie analytique, paraît triviale depuis que Descartes a montré comment associer les points de l'espace euclidien et les solutions numériques d'équations algébriques. Mais cette même représentation se révèle, quand il s'agit de mesurer la cardinalité des ensembles mis en jeu, extraordinairement difficile. La réponse à la question de la cardinalité du continu a mobilisé des efforts considérables et a favorisé l'apparition de nombreux concepts nouveaux[18]. Il ne serait pas possible de retracer ici tous ces développements mais on peut remarquer que l'un d'entre eux

---

[15] Il s'agissait au départ de savoir quelle est la cardinalité de l'ensemble formé par les points d'un intervalle de la droite réelle. Cette question, somme toute particulière, revêt cependant une grande importance dans la mesure où c'est elle qui permettrait de décider s'il est légitime de concevoir une échelle des cardinalités, maîtrisable par une arithmétique appropriée et dont l'opération fondamentale serait celle de l'opération de mise à la puissance, c'est-à-dire de multiplication infinie. Réussir à déterminer la taille du continu apparaît alors comme une stratégie indirecte en vue de faire avancer la réponse à la question de la détermination généralisée de la taille des ensembles. Cf. **M. Hallet**, *Cantorian set theory and limitation of size*, Clarendon Press, Oxford, 1984, p. 86 sq.

[16] C'est ainsi que Gödel commence son exposition du problème du continu cantorien : «Le problème du continu cantorien est simplement la question combien il y a de points sur une ligne droite dans l'espace euclidien. Une question équivalente est celle-ci : combien existe-t-il d'ensembles distincts d'entiers ?». **K. Gödel**, "What is Cantor's continuum problem ?", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, Vrin, Paris, 1992], pp. 509-531.

[17] Comme nous le verrons, il est possible de rejeter ce point de vue : Brouwer en particulier insistait sur le fait que le continu n'est pas formé d'un ensemble de points, comme le veut l'ensemblisme.

[18] Wang Hao constate combien le problème du continu a, en logique mathématique, favorisé les progrès au cours de ce siècle : «Le problème du continu est important et pose un défi pour des raisons variées. C'est d'abord une question naturelle, mais qui reste sans réponse depuis plus d'un siècle. Il traite des ensembles infinis les plus petits et des ordinaux infinis, parce qu'on peut le reformuler en se demandant s'il y a plus d'ensembles d'entiers que d'ordinaux dénombrables. Il a favorisé beaucoup de travail intéressant : en fait, les deux idées les plus intéressantes et les plus fécondes de l'axiomatisation de la théorie des ensembles, savoir la constructibilité et le forcing, furent toutes les deux introduites pour répondre partiellement à ce problème». **W. Hao**, *Popular lectures in mathematical logic*, Van Nostrand Reinhold Company, New York, 1981, p. 121.

a eu pour effet de “déréaliser” le continu [19]: puisque tout ensemble transfini admet un modèle dénombrable, les nombres réels, quel que soit le mystère qui entoure l’aspect presque arbitraire de leur cardinalité, ne devraient pas être considérés comme essentiellement différents d’autres ensembles numériques dont la cardinalité est dénombrable.

C’est habituellement en se plaçant du point de vue ensembliste de l’étude de la cardinalité du continu et en prenant en compte sa “déréalisation” que l’on critique le projet d’une étude de l’intelligence artificielle à partir de la notion de continu : le point de vue choisi dans cette étude serait finalement très arbitraire puisque l’opposition du discret et du continu ne serait pas aussi absolue qu’on le croit. Il n’y aurait pas lieu, en tout cas, d’en faire un angle d’attaque privilégié pour essayer de rendre compte du projet de l’intelligence artificielle. On doit répondre à cette objection en deux temps.

Premièrement, en se plaçant du point ensembliste, il n’est en effet pas question de revenir à une conception réaliste et naïve du continu des réels. Néanmoins, la possibilité d’exhiber un modèle dénombrable pour tout ensemble, y compris transfini, (c’est-à-dire, dans le cas qui nous occupe, y compris l’ensemble transfini dont la puissance est celle de l’ensemble des nombres réels), ne règle pas la question de la richesse du continu, puisque la construction d’un modèle dénombrable se fait seulement au prix d’une restriction axiomatique portant sur la notion d’ensemble[20]. Aussi l’objection n’a-t-elle aucune valeur absolue puisqu’elle met seulement au jour une insuffisance, propre à la méthode axiomatique, à caractériser adéquatement des classes infinies d’objets que l’intuition parvient cependant à circonscrire quand elle n’est pas bornée par un souci d’effectivité constructive.

---

[19] Cf. **T. Skolem**, “Sur la portée du Théorème de Löwenheim-Skolem” dans [**J. E. Fenstad ed.** *Collected Works*, Universitetsforlaget, Oslo, 1970], pp. 455-482.

[20] Cf. Les remarques de Bernays à la suite de l’exposé de Skolem dans **T. Skolem**, “Sur la portée du Théorème de Löwenheim-Skolem” dans [**J. E. Fenstad ed.** *Collected Works*, Universitetsforlaget, Oslo, 1970], pp. 455-482.

Deuxièmement, on peut faire remarquer qu'il y a d'autres points de vue mathématiques sur le continu que le point de vue ensembliste[21]. Ce n'est pas l'étude de ces points de vue en eux-mêmes qui nous retiendra. D'un point de vue général, c'est *l'intuition* mathématique du continu qui fait le fond de ce que nous appelons la question du continu et non pas seulement le problème de sa cardinalité. La question que nous posons n'est donc pas celle de savoir par rapport à quel modèle mathématique du continu le projet de l'intelligence artificielle s'est défini en tant que méthodologie essentiellement discrète. Le problème que nous voulons aborder est tout autre : si l'on part de l'hypothèse que la notion de continu entre dans les processus mentaux qui permettent la constitution d'une signification, il s'agit de savoir ce que devient la notion de continu dans le cadre du projet de l'intelligence artificielle, qui tente par le biais d'une méthodologie discrète de rendre compte de la constitution des représentations ainsi que de leur traitement.

412. Objection physique

La philosophie spontanée du chercheur en intelligence artificielle colporte un second présupposé concernant la notion de continu : celle-ci serait, du point de vue physique, une pure et simple *illusion*, car rien ne serait continu dans la nature[22]. Dès lors, l'absence de référence au continu ne serait pas un handicap

---

[21] Cf. **Salanskis J.-M.** et **Sinaceur H. dirs.**, *Le Labyrinthe du Continu*, Springer-Verlag, Paris, 1992 où les auteurs abordent cette question à la fois du point de vue historique et du point de vue des modèles mathématiques actuels.

[22] Cf. par exemple, **J.-P. Delahaye**, "Formulations mathématiques de la question: le monde est-il récursif ?", Cahiers du Crea n°15 : Méthodologie de la science empirique (2) sous la direction d'Alain Boyer, École Polytechnique, Paris, janvier 1993, p. 185-225 : «Pour donner un sens précis à la question "le monde est-il récursif ?" il faut avant tout accepter de se limiter à des modèles discrets du monde. Celui qui pense que le continu est essentiel dans la modélisation formelle du monde physique et que les discrétisations qu'on peut en faire (et qu'on fait nécessairement pour rendre applicable la physique) ne sont que des approximations insatisfaisantes par nature, ne franchira jamais ce premier pas. Notons, sur ce problème, que la nécessité dans laquelle nous sommes pour toute application d'un modèle physique de le discrétiser à un moment ou à un autre oblige les défenseurs de l'irréductibilité et de l'inévitabilité du continu à recourir à des affirmations ontologiques dont on imagine mal qu'elles soient prouvables et qui donc échappent à l'argumentation scientifique».

pour la méthodologie discrète de l'intelligence artificielle, puisque, en fin de compte, la notion de continu physique *n'aurait aucune réalité* [23].

Pour justifier cette assertion[24], on considère la réalité physique comme à la fois *ultime* et *discrète* : on s'appuie en particulier sur le fait que les électrons sont des parcelles discrètes de matière et que les échanges d'énergie à ce niveau électronique sont suffisants pour une théorie du calcul qui serait matériellement incarnée dans une machine[25]. Dès lors, l'intelligence artificielle n'a que faire du continu, réduit au rang d'illusion intuitive. Cette argumentation est réductionniste : elle considère que le niveau ultime de ce qui est appelé "nature physique" est de type granulaire et discret et qu'il suffit de parvenir à imiter ce niveau pour recueillir le fruit de toute discipline scientifique, à savoir son objectivité.

Mais il s'agit là d'un parti-pris et non pas seulement d'une prise de position relevant exclusivement de la physique[26] : le débat sur l'aspect continu ou discontinu de la *nature* physique engage en effet un débat philosophique sur ce qu'il faut attendre d'une *théorie* physique.

Il faut, pour comprendre les enjeux de ce débat, faire une remarque d'ordre historique : c'est au moment de la naissance de la mécanique quantique, à la fin des années 20, que les discussions sur l'aspect continu ou discontinu de la nature physique sont apparues sous leur forme actuelle[27]. La découverte de l'aspect discontinu de la matière à l'échelle atomique et subatomique semble

---

[23] C'est Von Neumann qui semble avoir popularisé cette idée, lui qui écrivait en 1932 que la naissance de la physique quantique était précisément liée à l'abandon du principe leibnizien de continuité : «l'existence du principe de continuité - *natura non facit saltus* - […] n'est au fond qu'une illusion». cité par **C. Chevalley**, "Continu et Discontinu dans la construction de la théorie quantique. Un exemple.", dans [*Le Labyrinthe du Continu*, **J.-M. Salanskis** et **H. Sinaceur dirs.**, op. cit.], p. 325.

[24] La conséquence majeure est que ce n'est pas l'ordinateur qui doit simuler la nature physique *mais bien l'inverse* : «La physique est essentiellement l'expression de la calculabilité du substrat spatio-temporel. Si l'on ne peut pas retrouver ces lois dans l'ordinateur, il faut alors en conclure que les lois de la physique sont erronées». **J. Ramunni**, *La physique du calcul; Histoire de l'ordinateur*, Hachette, Paris, 1989, p. 216.

[25] Plusieurs chercheurs, en intelligence artificielle et en informatique théorique, ont tenu devant moi ce raisonnement, en particulier Jacques Pitrat et Jean-Paul Delahaye. Il y a tout lieu de supposer qu'il est général tout en étant le plus souvent implicite.

[26] Tout le débat est intégralement rendu d'un point de vue historique dans **C. Chevalley**, "Continu et Discontinu dans la construction de la théorie quantique. Un exemple.", dans [*Le Labyrinthe du Continu*, **J.-M. Salanskis** et **H. Sinaceur dirs.**, op. cit.], pp. 327-337.

[27] Le débat remonte sûrement aux présocratiques, à Anaxagore en particulier.

remettre en question le cadre continuiste des théories physiques classiques dont la forme mathématique relève essentiellement de l'usage d'équations différentielles et partant, de l'analyse[28]. Cependant, deux théories, également admissibles d'un point de vue mathématique[29] tout en étant ontologiquement opposées d'un point de vue physique, sont à l'origine de ce qui allait devenir la mécanique quantique : la première, dûe à Heisenberg, est discontinue et utilise le formalisme mathématique du calcul des matrices; la seconde, dûe à Schrödinger, est continue et utilise le formalisme mathématique classique. Les deux points de vue se distinguent donc par le type de *réalité* physique que leurs formalismes mathématiques associent à leurs théories physiques[30].

L'enjeu du débat concernant l'aspect continu ou discontinu de la nature n'est donc pas de trancher définitivement la question de savoir si l'aspect continu ou discontinu l'emporte à un niveau ultime - comme le laisse supposer l'attitude réductionniste - mais bien de permettre une réflexion philosophique sur les *conditions physiques de constitution de l'objectivité*. Or ce ne sont pas les théories mathématiques elles-mêmes qui peuvent trancher la question de la constitution de l'objectivité, puisqu'elles sont, au moins dans cet exemple, formellement équivalentes. Il y a donc ici un débat proprement philosophique qui ne peut pas se régler en faisant seulement appel à des parti-pris réductionnistes.

On pourrait objecter qu'il n'y a peut-être rien de naturel dans notre appréhension intuitive de la réalité physique par le biais de modèles continus. Il est en effet fort possible que les progrès de l'analyse depuis l'avènement du calcul infinitésimal aient modelé notre intuition de telle sorte que l'appel à des modèles continus en physique nous paraisse, de façon erronée, tout naturel. Il se pourrait qu'une réforme de nos modes d'intuition, bien qu'elle ne se commande

---

28 Cf. **A. Messiah**, *Traité de mécanique quantique,* 2ème édition, Dunod, Paris, 1969, chapitre 1.

29 J. von Neumann a montré qu'elles étaient deux représentations équivalentes du point de vue des espaces de Hilbert. Cf. **C. Chevalley**, "Continu et Discontinu dans la construction de la théorie quantique. Un exemple.", dans [**J.-M. Salanskis** et **H. Sinaceur dirs.**, *Le Labyrinthe du Continu*, op. cit.], p. 329.

30 Comme le remarque C. Chevalley : «L'opposition entre le continu et le discontinu, […] dérivait donc d'une opposition entre différents processus de construction de la réalité physique, différents modes de constitution de l'objectivité». **C. Chevalley**, "Continu et Discontinu dans la construction de la théorie quantique. Un exemple.", dans [**J.-M. Salanskis** et **H. Sinaceur dirs.**, *Le Labyrinthe du Continu*, op. cit.], p. 331.

pas, soit en cours et ce, par le biais de modèles discrets[31]. Si tel était bien le cas, il faudrait remettre en chantier la question philosophique de la constitution de l'objectivité physique. Mais le débat paraît encore trop récent pour que, à l'aide de résultats physiques tangibles, on puisse déjà en mesurer philosophiquement les tenants et les aboutissants[32].

Pour nous, dont la tâche consiste à réfléchir à la constitution de l'objectivité dans le domaine de l'intelligence artificielle, une remarque s'impose : tant que le continu, *illusoire ou pas*, apparaîtra comme inaccessible à la méthodologie discrète de l'intelligence artificielle, il y aura quelque chose relevant à la fois du domaine intuitif et nécessaire, jusqu'à nouvel ordre, à l'élaboration de la physique, qui restera étranger au but fondationnel que se propose de réaliser l'intelligence artificielle. Car enfin, si le continu est une illusion, encore faut-il être en mesure de le montrer en exposant le mécanisme qui engendre celle-ci. Or de ce mécanisme, il n'est *nulle part* question chez ceux qui considèrent que la notion de continu n'est qu'une apparence.

De façon moins négative, on peut tirer une leçon philosophique du débat moderne concernant la constitution de l'objectivité physique. On remarque en effet que l'opposition du continu et du discontinu dans une théorie physique pose le problème du statut à accorder à l'intuition et à la langue naturelle dans la constitution de l'objectivité[33]. Pour l'interprétation continuiste, notre intuition comme notre langue nous poussent à construire une représentation continue du monde. Pour l'interprétation discontinuiste, c'est au contraire en délimitant mathématiquement une région de l'observable que l'on parvient à éliminer les intuitions et les mots, toujours trompeurs. Le débat tourne donc autour de la

---

[31] On trouve trace d'une telle idée dans le compte-rendu du livre de **W. H. Zurek**, *Complexity, Entropy and the Physics*, Addison-Wesley Publishing Company, 1990, par J.-P. Delahaye dans la Lettre de l'Institut Henri Poincaré, n°5, novembre 1991.

[32] Il me semble peu probable cependant qu'une telle recherche puisse ressusciter dans sa forme originelle le programme de Carnap, que plus personne aujourd'hui ne semble prêt à poursuivre.

[33] Niels Bohr écrit à Schrödinger : «(…) la définition de tout concept ou plutôt de tout mot présuppose la continuité des phénomènes et par suite devient ambiguë dès que cette présupposition fait défaut». cité dans **C. Chevalley**, "Continu et Discontinu dans la construction de la théorie quantique. Un exemple.", dans [*Le Labyrinthe du Continu*, **J.-M. Salanskis** et **H. Sinaceur dirs.**, op. cit.], p. 335.

confiance que l'on peut accorder à ce qui se présente naturellement à l'intuition, à savoir les syntagmes de la langue naturelle et les concepts qui en dérivent.

Ce qui nous retiendra donc ici d'un point de vue général, c'est que le débat au sujet de la constitution d'une objectivité physique en vient à poser la question du rôle de la langue dans son rapport aux concepts. Or ce débat est précisément celui que nous avions déjà rencontré tout à l'heure dans le contexte du continu mathématique puisqu'il était question de la confiance qu'il fallait accorder aux traductions linguistiques formalisées d'actes intuitifs de pensée. Nous tenterons de montrer, au cours de ce travail, comment une psychologie rationnelle exige l'élaboration d'un concept du continu du point de vue de la *constitution de toute expression linguistique* : ce serait *l'expression* des concepts dans un langage qui nécessiterait l'intervention du continu dans une psychologie rationnelle. Le continu serait alors interprété comme nécessairement présupposé pour parvenir à exprimer des concepts qui sans lui, resteraient prisonniers de l'infra-langagier.

Il ne s'agit pas ici de faire porter notre attention sur le langage en tant que tel: une attitude tatillonne à l'égard du langage paraît être le meilleur moyen de ne pas saisir ce qui constitue la difficulté qui nous occupera tout au long de cette étude, à savoir l'écart qui sépare les choses de notre capacité à les exprimer. Nous ferons donc l'hypothèse qu'il n'y a pas intrinsèquement de problème philosophique du langage mais qu'il y a en revanche un problème philosophique de *l'accès* au langage, c'est-à-dire des moyens cognitifs mis en œuvre pour parvenir à exprimer ce qui ne relève pas, ou pas encore, de l'expression[34].

42. Objections phénoménologiques

---

[34] Nous suivons ici les remarques de **W. Hao**, *From mathematics to philosophy*, Routledge and Kegan Paul, London, 1974, p. 14 : «Quand la syntaxe et le *Tractatus* se furent tous deux révélés inadéquats, plutôt que de se départir ou de dépasser la logique et le langage, on essaya d'amplifier les constructions logiques et d'affiner l'analyse linguistique. Le résultat fut que nous avons dérivé vers des discours opaques sur la sémantique et sur des considérations empiriques et bancales de pragmatique et de linguistique. (…). Mais si l'on se rappelle que l'infatuation première pour le langage fit l'objet de fausses promesses, alors il n'est plus nécessaire de considérer comme sacrée la préoccupation du philosophe pour le langage».

Les objections de nature phénoménologique touchent moins le sens même de notre projet que la méthode employée.

421. Le rôle du corps

On sait que le projet de l'intelligence artificielle a été critiqué du point de vue de la phénoménologie. Hubert Dreyfus en particulier s'est lancé depuis le milieu des années soixante dans une vaste critique des buts que se propose d'atteindre l'intelligence artificielle[35]. Cette critique est radicale dans la mesure où, en se plaçant d'un point de vue phénoménologique, elle ne reconnaît pas dans l'intelligence artificielle un projet véritablement scientifique.

Mais il nous semble que toute objection directement phénoménologique est inopérante à l'égard de la position de l'intelligence artificielle, parce que le point de vue de l'intelligence artificielle ne reconnaît comme pertinent que le cadre calculatoire qui lui sert de fondement : toute considération sortant de ce cadre, et en particulier toute critique, apparaît donc comme manquant l'objet qu'elle vise. C'est la raison pour laquelle, jusqu'à présent, aucune objection d'inspiration phénoménologique adressée à l'intelligence artificielle n'a eu d'incidence réelle sur la conception même du projet : en particulier, toute critique de l'intelligence artificielle par le biais de questions touchant à l'aspect incarné du corps ne peut pas être entendue, parce que ces questions sortent *a priori* des limites algorithmiques que s'est fixée l'intelligence artificielle. Habituellement, chacun reste campé sur ses positions, et on en reste à l'exposition d'articles de foi sur ce que l'intelligence artificielle est censée "pouvoir" ou "ne pas pouvoir" faire[36].

Dans cette situation, pour réussir à introduire l'idée de la place du corps incarné, il faut étudier non pas le corps en lui-même mais les schèmes abstraits de pensée d'origine corporelle qui s'expriment à un niveau linguistique et logique.

---

[35] Cf. par exemple, **H. Dreyfus**, *What Computers Can't Do*, Harper and Row, New York, 1972 et 1979, traduction française *Intelligence Artificielle: mythes et limites*, Flammarion, Paris, 1984, pp. 433-439.

[36] On en a un exemple dans l'appendice à la traduction française du livre de **H. Dreyfus**, *Intelligence Artificielle: mythes et limites*, op. cit., pp. 433-439.

La notion de continu joue ici un rôle tout à fait spécifique : c'est par un retour aux schèmes corporels qui soutiennent la façon dont le continu peut être représenté qu'une réflexion sur la place du corps peut, nous semble-t-il, devenir légitime. De ce point de vue, la notion de machine, apparemment essentiellement logique, jouera, dans notre étude, un rôle capital: la notion de machine de Turing est en effet sous-tendue par un schème corporel dont il faudra réussir à montrer la prégnance jusque dans le domaine logique. Le rôle de la notion logique de machine de Turing dans la constitution des représentations permettra d'introduire une réflexion sur son rapport à la notion de continuité. L'étude du concept de continu nous permet ainsi d'adopter un point de vue qui ne soit pas essentiellement négatif à l'égard du projet de l'intelligence artificielle et qui autorise cependant à en faire une critique raisonnée[37], précisément parce qu'il permet d'introduire une critique de nature phénoménologique à partir du point de vue algorithmique, seul reconnu par l'intelligence artificielle.

422. Le rôle du langage

On peut faire une remarque presque stylistique sur le projet de l'intelligence artificielle. Comme tout travail scientifique, le projet en question véhicule un certain nombre de métaphores. Nous nous tiendrons, à ce sujet, au principe suivant : on ne doit pas considérer que ce qui est métaphorique sort nécessairement des limites de la science. Comme le faisait remarquer Michael Arbib dès le titre de l'un de ses ouvrages[38], le projet de l'intelligence artificielle

---

[37] Le mot "critique" est ambigu. Il y a tout d'abord une critique au sens courant : critique des slogans claironnants tels que "intelligence artificielle" ou "penser, c'est calculer", nécessaires sans doute pour obtenir des crédits d'institutions crédules mais qu'il est à peine besoin de fustiger d'un point de vue philosophique, à moins de ressentir le besoin de faire une critique philosophique de la publicité. Mais il y a un second sens au mot critique, de nature plus relevée. Il s'agit d'une critique au sens kantien, c'est-à-dire d'une analyse raisonnée visant à mettre en lumière les conditions d'objectivité d'un domaine du savoir. C'est dans ce second sens que nous entendons le mot de critique, même si notre argumentation n'est pas, dans le fond, de nature kantienne. L'hypothèse que nous faisons est donc que l'expression d'intelligence artificielle recouvre autre chose qu'un slogan publicitaire bien que, comme dans tout domaine où la technologie est en jeu, de gros capitaux soient engagés et donc aussi des intérêts qui n'ont rien à voir avec la connaissance.

[38] **M. Arbib**, *The Metaphorical Brain; An Introduction to Cybernetics as Artificial Intelligence and Brain Theory*, Wiley-Interscience, New York, 1972, Préface, p. vii : «Notre but est de faire

s’est constitué autour de deux métaphores : d’une part, les êtres humains sont des machines; d’autre part, les êtres humains sont des animaux. Nous prendrons au sérieux ces deux métaphores, mécanique et biologique, sans chercher à les édulcorer et en tentant d’analyser leurs implications philosophiques, en particulier dans les textes de Turing auxquels nous aurons affaire. Aussi, comme nous l’avons déjà précisé, sans focaliser notre étude sur des questions linguistiques, nous tenterons d’accorder une place à l’expression telle qu’elle se donne, dans la profusion des métaphores. C’est à ce prix, qui pourrait sembler excessif à qui aurait une conception étroite de la rationalité de la science, qu’une interprétation philosophique des fondements de l’intelligence artificielle nous paraît possible.

---

parvenir à une compréhension du cerveau par le biais de deux métaphores : la métaphore cybernétique, “Les êtres humains sont des machines”, et la métaphore évolutionniste, “Les êtres humains sont des animaux”».

Première partie

La psychologie dans la logique

«Dans ce but nous analysons la science, non pas pour en extraire ce qu'on a considéré comme ses *résultats* […] et moins encore pour nous inspirer de ses méthodes […], mais en la considérant comme la matière brute du travail, comme un spécimen saisissable de la pensée humaine et de son développement».

Émile Meyerson, *Identité et réalité*, p. VIII

Introduction

Nous allons essayer, dans cette partie, de mettre en rapport deux questions qui, au premier abord, semblent ne rien à avoir en commun : la première est d'ordre mathématique et porte sur la notion de calcul; la seconde est d'ordre psychologique et porte sur la possibilité d'une représentation de la pensée. Nous essayerons de montrer que c'est du rapprochement des deux notions de calcul et de représentation, qui appartenaient auparavant à des savoirs différents, qu'est née l'idée d'une intelligence artificielle.

1. Le rapport des mathématiques et de la psychologie

Quand on s'interroge sur la nature de la notion de calcul en mathématique, on s'aperçoit qu'elle procède d'une attitude constructive. A l'origine, la notion de construction désignait la manipulation, par des procédés abstraits, des entités géométriques construites à la règle et au compas. Le sens s'est peu à peu déplacé et désigne tout procédé effectué en un nombre fini d'étapes permettant la manipulation d'entités abstraites, figures ou signes : la notion s'apparente à celle d'algorithme. La caractérisation de la notion d'algorithme a fait l'objet, au cours du XXème siècle, d'un certain nombre de thématisations[39]. Elles sont apparentées dans la mesure où elles tentent de préciser ce que l'on entend par calcul quand on se limite à la manipulation, par des procédés effectués en un nombre fini d'étapes,

---

[39] Comme le fait remarquer J. Largeault au sujet de la pratique mathématique des siècles passés : «Le constructivisme existait en mode dispersé; il ne reçut une impulsion que lors de l'extension des méthodes abstraites contemporaines de la théorie des ensembles, et finit sous nos yeux par se centrer autour des notions de règles et d'algorithme». **J. Largeault**, *L'intuitionisme*, Coll. Que Sais-je ? n° 2684, Presses Universitaires de France, Paris, 1992, p. 110. Les formes du constructivisme sont multiples; on en compte au moins trois grands courants qui correspondent à trois périodes chronologiques : l'école appelée rétrospectivement “pré-intuitionniste” (Borel, Lebesgue); l'école intuitionniste (Brouwer, Weyl); l'école d'analyse non-standard (Robinson, Reeb, Nelson). Turing a été l'instigateur d'un nouveau courant, qui a donné naissance à l'algorithmique et à l'informatique théorique.

d'entités dénombrables, qu'elles soient en nombre fini ou infini.

Une de ces thématisations, d'ordre mathématico-logique[40], a été effectuée par Turing. Celui-ci a proposé une façon d'envisager la question de la nature du calcul en s'aidant d'une notion mathématique qu'il invente et qui s'apparente à celle de "machine à calculer". Cette thématisation a un statut exemplaire quand on se place, comme nous le faisons, d'un point de vue psychologique, parce que Turing fait appel, pour introduire sa nouvelle notion de "machine à calculer", à la figure du mathématicien calculant et aux états de son esprit, contrairement aux autres thématisations de la notion de calcul qui ne font référence, pour décrire la notion de calcul, qu'à des signes et à des compositions de signes selon des opérations. La formulation de Turing, de par son caractère psychologique, apparaît ainsi de nature plus réflexive que les autres, dans la mesure où elle rapporte à la capacité de penser l'introduction de ce qui apparaît comme un nouveau concept de nature mathématique. C'est cette apparition presque inédite[41]

---

[40] Par "mathématico-logique", on fait référence au point de vue que Brouwer qualifiait de "logico-linguistique". Parlant du logicisme en mathématique en le désignant sous le nom "d'ancienne école formaliste", Brouwer décrivait ainsi l'attitude mathématique propre à ce courant : «Encouragée par l'importance du rôle joué par la méthode logico-linguistique en géométrie, l'ancienne école formaliste plongea la logique et les mathématiques dans une science linguistique unique travaillant sur des mots ou des symboles dénués de sens en employant des lois logiques, ce qui efface la différence de caractère entre logique et mathématique et les dépouille de leur autonomie». **Brouwer L. E. J.,** "Points and Spaces", *Collected Works*, I, p. 523, traduction française dans **J. Largeault**, *L'intuitionisme*, Coll. Que Sais-je ? n° 2684, Presses Universitaires de France, Paris, 1992, p. 109.

[41] A notre connaissance, seuls Bolzano, Cantor et Dedekind avaient, avant Turing, justifié l'introduction d'un concept mathématique - les notions d'ensemble infini et de cardinalité - par le biais de considérations psychologiques. Dans "Was ist und was sollen die Zahlen ?", Dedekind justifie ainsi l'introduction de la notion d'ensemble infini : « §66. Théorème : il existe des systèmes infinis. Démonstration (une considération semblable se trouve au § 13 des *Paradoxes de l'infini* de Bolzano, Leipzig, 1851) : le domaine de mes pensées, c'est-à-dire l'ensemble S de toutes les choses qui peuvent être objets de ma pensée est infini. Car si a désigne un élément de S, la pensée s', que s peut être objet de ma pensée, est elle-même un élément de S. Si l'on considère le même élément comme l'image ϕ (s) de l'élément s, l'application ϕ de S ainsi définie a la propriété que l'image S' est une partie propre de S; et en particulier, S' est une partie propre de S, car il y a dans S des éléments (par exemple mon moi) qui sont distincts de toute pensée de s' de ce genre et pour cette raison ne sont pas contenus dans S'. Enfin, il est clair que si a, b, sont des éléments différents de S, leurs images a', b' sont aussi différents, donc que l'application est une application distincte. Par conséquent, S est infini. C.Q.F.D.». **Dedekind R.,** *Continuité et nombres irrationnels*, trad. franç. par J. Milner revue par H. Sinaceur, Les cahiers d'Ornicar, Paris 1978. Chez Cantor, la définition du nombre cardinal est justifiée par un processus psychologique qu'il appelle l'abstraction; Cf. **Cantor G.,** "Mitteilungen zur Lehre vom Transfiniten", section VIII, pp. 411-412, cité dans **Dauben J. W.,** *Georg Cantor, His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, Mass., 1979, p. 221. A ces trois auteurs, on peut sans doute

de la psychologie sur le terrain des mathématiques qui devra nous retenir au premier chef.

La seconde question présente un aspect directement philosophique : comment réussir à penser la pensée ? On sait que cette question revient à se demander comment opérer une rupture dans l'adhérence de la pensée à elle-même, rupture permettant de s'en faire une représentation. On a vu que seule une expérience indirecte reposant sur l'usage de signes semble permettre d'opérer cette rupture dans la continuité. R. Ruyer donne l'exemple d'une expérience indirecte concernant notre faculté de vision[42]:

> **«[…] il nous arrive d'avoir à obtenir un jugement sur notre propre champ visuel, soit par l'intermédiaire du spécialiste auquel nous faisons rapport, soit par nous-mêmes en tant qu'oculiste improvisé, en cherchant à voir des objets quelconques : "Je puis lire sur le tableau A H J X …" Cette auto-observation n'est pas "un regard jeté sur …" notre vision. C'est une constatation indirecte de succès ou d'échec relativement à une réussite idéale ou normale.»**

Deux remarques doivent être faites concernant l'aspect indirect de l'expérience nécessaire pour rompre la continuité de la pensée avec elle-même.

Premièrement, dans l'exemple décrit par R. Ruyer, c'est l'utilisation d'objets comme symboles qui rend l'expérience indirecte : dans le cas de la visite chez l'oculiste, l'expérience de la lecture des lettres (ou de la description précise de tout autre objet) permet de projeter sur un plan externe, dans le monde des objets, les caractéristiques propres à la faculté de la vision dans son usage interne. On crée donc artificiellement un intermédiaire, qui permet de projeter ce qui se donne sans intermédiaire, la vision. Ce qui est *signe* est précisément ce qui relève de cette projection : objet extérieur (en l'occurrence, des lettres) il est investi d'un rôle de *représentation* puisque sa perception (la vision des lettres) n'est pas opérée pour elle-même mais pour ce qu'elle signifie de la vision elle-même. La

---

ajouter l'introduction par Brouwer de la notion de suite de choix libre comme acte de production du sujet créateur, si tant est qu'il s'agisse encore d'une psychologie au sens que l'on donne habituellement à ce terme. Cf. par exemple, **Brouwer L. E. J.**, "Base historique, principes et méthodes de l'intuitionisme", traduction française dans [**Largeault J.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 445-458.

[42] **Ruyer R.**, *Paradoxes de la conscience et limites de l'automatisme*, Albin Michel, Paris, 1966, p. 24.

signification est donc artificiellement déplacée des objets qu'elle vise aux facultés qui la rendent possible : elle est détournée de son usage direct qui consiste, dans le cas de la vision des lettres, à établir des rapports entre les lettres et les mots pour être indirectement associée à ses conditions de production. La notion de représentation peut être définie, dans une première approche, comme le résultat de ce déplacement.

Deuxièmement, l'exemple de R. Ruyer tend à montrer que pour réussir à circonscrire le pouvoir d'une faculté psychologique (en l'occurrence, la vision), il faut réussir à se placer d'un point de vue dans lequel cette limitation ne joue pas : ici, c'est la personne de l'oculiste qui, en occupant une position suffisamment proche du tableau où sont inscrites les lettres, n'est pas soumise à une limitation dans sa vision, au moins pas à l'échelle de grandeur où s'opère le déchiffrement des lettres. Dans le cas de l'appréhension par un individu de sa propre pensée, la nécessité de s'en remettre à autrui pour tâcher d'en circonscrire les bords est tout aussi présente. On en conclut, d'un point de vue général, qu'il doit exister pour chaque individu des pensées inaccessibles à sa pensée, celles précisément qui résistent à sa compréhension[43]. C'est seulement dans le temps qu'une prise de conscience de l'existence de pensées non directement accessibles peut avoir lieu et une fois que cette limitation a été localement dépassée grâce à l'aide d'autrui.

L'accès au langage et à la temporalité sont donc les moyens par lesquels la pensée peut prendre une distance par rapport à elle-même. C'est ce qui rend possible la constitution d'une théorie de la psychologie, qui doit nécessairement délimiter un champ spécifique dans lequel des phénomènes vont pouvoir être étudiés. La différentiation de la pensée d'avec elle-même apparaît alors comme un processus qui n'a pas nécessairement de fin; elle se constitue en effet dans une ouverture temporelle indéfinie et dans une expérience seulement indirecte d'elle-

---

[43] Comme le fait remarquer Fodor : «Et, dès lors que l'on admet que l'esprit a une structure endogène, on voit mal comment la théorie pourrait ne pas imposer de limitations à la classe des croyances accessibles. Ces remarques sont tout à fait indépendantes de la question de la modularité; elles suggèrent qu'en un sens *toute* théorie de l'esprit doit accepter le fait que celui-ci a un domaine propre. La seule question épistémiquement intéressante est donc de savoir s'il est probable que certaines pensées inaccessibles soient à la fois intéressantes et vraies.» **Fodor J. A.**, T*he Modularity of Mind*, MIT Press, 1983, trad. franç. *La modularité de l'esprit*, Minuit, Paris, 1986, p. 161.

même menée grâce à la maîtrise de signes dans des représentations.

C'est par le biais de moyens symboliques qu'il est possible à la pensée d'opérer un retour sur elle-même : ainsi, même l'absence originelle de délimitation de la pensée devient descriptible dans la sphère symbolique quand on fait usage d'une notion telle que le continu. Cette notion apparaît comme assurant le passage entre l'aspect indifférencié et non-linguistique de la pensée dans son rapport originel à elle-même et les états bien différenciés de la pensée, susceptibles d'être décrits comme des éléments discrets : entre la continuité de la pensée et son modèle qui la décrit comme relevant d'un fonctionnement discret, la notion de continu assure une transition.

Si on peut concevoir les éléments de la pensée sur le mode du fonctionnement discret, on voit cependant que cette description n'est possible que par l'appel à un état premier non différencié. Ainsi l'idée d'un fonctionnement de la pensée demande-t-elle à être explicitée dans une démarche temporelle, à partir d'un état non-différencié. Si l'on considère l'intelligence artificielle comme une théorie du fonctionnement de la pensée, alors c'est à partir d'un état non-différencié que l'on caractérisera par la notion de continu qu'il faut réussir à en penser l'émergence.

Or c'est précisément ce à quoi Turing aboutit quand il invente son concept de “machine à calculer” en se servant de l'introspection pour décrire les états mentaux du mathématicien au travail : c'est bien en effet pour résoudre la question de la calculabilité des nombres réels, c'est-à-dire des nombres qui forment le continu mathématique, qu'il invente son concept de “machine à calculer”. Cette conjonction de faits n'est pas le fruit d'une coïncidence : il doit y avoir un rapport profond - qu'il nous reste à expliciter - entre l'appel à l'introspection, le calcul des réels et la caractérisation de la pensée comme fonctionnant sur un mode discret.

On voit maintenant pourquoi on peut tenter de mettre en rapport les deux questions que nous venons de mentionner, l'une mathématique et l'autre philosophique. Dans la première, un concept mathématico-logique, celui de calcul, ne s'explicite que par un appel aux facultés du sujet. Dans la seconde, la pensée ne devient accessible à elle-même que par le biais d'une identification à un fonctionnement de type discret. Bref, dans la première question, pour penser cet objet mathématique particulier qu'est le calcul, on fait appel à la notion subjective de pensée tandis que dans la seconde, pour penser la pensée sur le mode de l'objet, on fait appel à la notion de fonctionnement. C'est ce double mouvement qu'il va falloir tenter d'analyser, parce que c'est lui qui caractérise en propre ce que l'on nomme "l'intelligence artificielle".

On peut s'aider, pour concevoir cette question, de la tradition philosophique.

2. Un exemple de la tradition : le *Théétète*

Un exemple célèbre de la tradition et qui remonte à Platon, opère le rapprochement que l'on vient de décrire : dans le *Théétète*, Platon tente de caractériser, en énonçant un certain nombre d'hypothèses, la démarche de la pensée[44]. Or ces hypothèses ne sont formulées par Platon qu'une fois qu'il a d'abord décrit la constitution d'une nouvelle opération mathématique[45], l'opération racine. Ainsi la méditation mathématique sur l'universalité de l'opération racine et le type de nombres auquel l'opération permet l'accès, les nombres irrationnels - qui font partie de ce que l'on appelle aujourd'hui les nombres réels[46] -, conduit Platon à s'interroger sur le problème philosophique du modèle à employer pour répondre à la question : comment en est-on venu à penser cette pensée ?, et dans une interprétation plus générale et plus libre à l'égard de Platon : en quoi le continu mathématique des nombres réels est-il partie prenante dans la possibilité d'une réflexion de la pensée sur elle-même ?

Faisons une hypothèse philosophique : le cas particulier décrit dans le *Théétète* est *exemplaire* par le rapport qu'il parvient à instaurer entre un questionnement d'ordre mathématique et un questionnement d'ordre

---

[44] **Platon**, *Théétète*, 197 a-b où Platon compare d'une part la pensée à un colombier et les pensées à des colombes et d'autre part la pensée à un morceau de cire sur lequel les pensées viennent se marquer.

[45] **Platon**, *Théétète*, 147 d-148 e.

[46] Cf. **J.-L. Gardies**, *L'héritage épistémologique d'Eudoxe de Cnide*, un essai de reconstitution, Vrin, Paris, 1988, p. 69.

philosophique. Plus précisément, l'exemplarité du *Théétète* vient du rapport instauré entre la découverte d'un type de nombres auxquels l'opération racine permet l'accès, les nombres que l'on qualifierait aujourd'hui de réels et la découverte d'un questionnement sur la nature de la pensée. L'hypothèse que nous faisons est donc la suivante : la découverte d'une méthode mathématique permettant de cerner la nature des nombres réels, et par ce biais de rendre compte du continu, a des incidences sur l'élaboration philosophique d'un modèle pour penser la pensée[47]. Il y aurait donc un rapport, qu'il faudra tenter d'éclaircir, entre d'une part la détermination mathématique du continu par le biais d'une méthode et d'autre part la construction de modèles pour penser la pensée. Une détermination relevant de la technique mathématique aurait ainsi à voir avec une détermination proprement philosophique.

Ainsi trois pôles se dégagent-ils du rapprochement effectué entre les deux questions : la constitution d'une nouvelle *méthode* mathématique permettant de décrire tout ou partie du champ numérique rendant compte du *continu* a pour conséquence l'élaboration philosophique de *modèles* pour penser la pensée. C'est en gardant en mémoire ces trois pôles qu'il est possible de comprendre le rapprochement entre une question de nature mathématique et une autre, de nature philosophique.

C'est ce mouvement, au premier abord mystérieux, que nous voudrions décrire dans un contexte différent de celui de Platon mais qui, somme toute, lui est apparenté. Il est en effet possible de décrire la constitution du concept de "machine de Turing" ainsi que ses implications philosophiques de la même manière que pour l'exemple de l'opération racine dans le *Théétète* : selon notre hypothèse, ce serait la constitution d'un nouveau concept mathématique

---

47 Au XX$^{\text{ème}}$ siècle, c'est sans doute le paradoxe de Richard qui joue le rôle de point de départ. En partant de l'hypothèse que les nombres réels définissables en un nombre fini de mots forment un ensemble dénombrable, on peut grâce à l'utilisation de l'argument de diagonalisation dû à Cantor, parvenir à une contradiction. Cf. **J. Richard**, "Les principes des mathématiques et le problème des ensembles", (1905), Revue générale des sciences pures et appliquées, 16, 541, reprint dans [*Logique et Fondement des mathématiques*, Anthologie (1850-1914), **F. Rivenc et P. de Rouilhan eds.**, Payot, Paris, 1992], pp. 271-275. J. Mosconi, dans la présentation du texte, insiste sur l'importance du paradoxe pour les «grands résultats négatifs des années 30 (Gödel, Tarski, Turing, …) et leurs prolongements».

permettant de déterminer la classe des *réels calculables*, qui aurait des conséquences philosophiques sur la constitution possible de modèles pour penser la pensée, modèles qui relèvent de ce que l'on a coutume d'appeler "l'intelligence artificielle". Nous retrouvons dans ce contexte les trois pôles que nous évoquions à l'instant dans l'exemple du *Théétète* et dont il va être question plus avant dans les pages qui suivent.

Il faut donc garder en mémoire, au cours de cette partie qui expose la nature de la notion de calcul principalement du point de vue du concept de machine de Turing, que le but est de parvenir à reproduire le mouvement de pensée accompli par Platon dans le *Théétète* et, par ce biais, de parvenir à se poser la question de la légitimité philosophique d'un certain nombre de modèles psychologiques visant à décrire la pensée sur le mode du fonctionnement.

Chapitre I

Ce que l'intelligence artificielle doit au débat sur les fondements des mathématiques

La philosophie dite "de l'esprit" qui émane de la réflexion sur l'intelligence artificielle appartient à ce qu'il est convenu d'appeler la philosophie analytique. Aussi la philosophie de l'esprit hérite-t-elle de l'aspect logiciste de la philosophie analytique, au sens que ce terme revêt dans le débat sur les fondements des mathématiques depuis le début du XXème siècle[48]. Pourtant, ni la notion de calcul, ni son rapprochement avec celle de représentation ne doivent grand-chose au logicisme. En particulier, le rapport que nous allons étudier dans cette partie entre la notion logique de calcul et celle, psychologique, de représentation, s'est constitué non pas dans un cadre de pensée logiciste mais dans le débat entre les tenants du formalisme et ceux de l'intuitionisme portant sur la valeur de l'idée de formalisation en mathématique. Il y a donc ici un décalage entre d'une part la réflexion philosophique et d'autre part l'apparition historique du problème épistémologique de la notion de calcul dans son rapport à la notion

[48] Comme le remarque D. Andler : «Aujourd'hui nommée "philosophie analytique", ce n'est pas une doctrine, mais un mode ou une méthode, une façon de philosopher. Cette philosophie de l'esprit, plus encore peut-être que les origines logico-épistémologiques de ses principales thématiques, explique qu'aujourd'hui ce soit elle qui contribue, plus que tout autre, aux sciences cognitives (…)». **D. Andler**, "Calcul et représentations : les sources", dans [**Andler D. dir.**, *Introduction aux sciences cognitives*, Folio Essais, Gallimard, Paris, 1992], pp. 19-20.

de représentation[49]. Ce décalage nous paraît dommageable à une bonne compréhension de la constitution de la problématique de l'intelligence artificielle.

Il faut donc commencer par essayer de rétablir, avant d'en venir à l'étude de la notion de calcul proprement dite, ce que le débat sur l'intelligence artificielle doit aux trois réflexions épistémologiques sur les fondements des mathématiques, le logicisme, le formalisme et l'intuitionisme[50].

1. Ce que l'intelligence artificielle doit au logicisme

C'est la thèse logiciste en philosophie des mathématiques que retient habituellement le philosophe qui se penche sur les fondements de l'intelligence artificielle, parce qu'elle constitue d'emblée sa philosophie spontanée[51]. De ce point de vue, les travaux logiques et philosophiques de Frege et de Russell dominent de façon plus ou moins consciente les esprits.

11. La notion frégéenne de système formel

L'intelligence artificielle semble avoir directement hérité de la notion de *système formel*, telle qu'elle a été élaborée par Frege dès 1879 en vue de rendre compte, d'un point de vue entièrement logique, de la nature de l'arithmétique et plus généralement de la nature des mathématiques[52]. Par système formel, on entend un système axiomatique développé dans une langue entièrement formalisée et apte à

---

[49] Ce décalage s'explique essentiellement pour des raisons de contingence historique : la réflexion philosophique sur l'intelligence artificielle est née dans une tradition de philosophie analytique de tendance logiciste à partir des années soixante de ce siècle, aux États-Unis.

[50] G. Kreisel remarque que le débat entre partisans et opposants de l'intelligence artificielle reproduit le débat qui a eu lieu il y a un siècle entre formalistes et antiformalistes. Selon lui, le débat philosophique sur l'intelligence artificielle «tirerait parti d'une référence explicite à l'indubitable progrès logique […], qui a eu lieu depuis les discussions, il y a un siècle, entre les (mêmes) partisans et opposants qui étaient alors appelés "formalistes" et "antiformalistes"». Ce «progrès logique» sur la question du formalisme consiste principalement pour Kreisel en la mise au jour de l'idéal de "rigueur informelle" telle qu'elle est contenue dans la thèse de Church-Turing. Nous aborderons l'analyse de cette thèse dans le chapitre suivant. Cf. **G. Kreisel**, "Church's Thesis and the Ideal of Informal Rigour", Notre Dame Journal of Symbolic Logic, 28, 4, October 1987, p. 516.

[51] Cf par exemple **Z. Pylyshyn**, *Computation and Cognition, Toward a Foundation for Cognitive Science,* A Bradford Book, MIT Press, Cambridge, Massachusetts, 1984, p. 198.

[52] Cette démarche est celle de Frege dans *Begriffschrift*, traduction française partielle dans [**F. Rivenc** et **P. de Rouilhan dirs.**, *Logique et Fondements des mathématiques*, Anthologie (1850-1914), Payot, Paris, 1992], pp. 93-129.

rendre compte du sens des notions mathématiques élaborées par ailleurs. Cette approche de la réalité mathématique possède deux traits spécifiques : d'une part, la logique est considérée comme une théorie *universelle* de la quantification, c'est-à-dire une théorie dans laquelle tous les objets envisagés sont traités de la même manière, sans considérations domaniales; d'autre part, il n'y a aucun sens pour les notions mathématiques envisagées en dehors du système formel : en particulier, il n'y a pas de questions méta-systématiques. Ainsi, même si le système formel possède des règles logiques qui lui sont nécessairement extérieures[53], ces règles n'ont pas de contenu logique en dehors de l'usage qui en est fait dans le système formel.

L'intelligence artificielle, de façon plus ou moins explicite, hérite de cette attitude fondationnelle : on se sert aussi dans ce domaine de la notion de système formel pour définir ce que l'on entend par traitement calculatoire des symboles. On tient alors un raisonnement de type analogique à propos de la nature de la pensée : si la notion de système formel joue le rôle de cadre *a priori* permettant de définir l'objectivité mathématique en général, la notion de calcul interprétée comme fondement d'une "science de l'esprit" permet de considérer l'esprit lui-même comme un système formel. La phrase de Hobbes selon laquelle "penser n'est rien d'autre que calculer"[54] et qui, comme telle, n'a guère de sens sinon polémique, reçoit alors un contenu précis : penser, c'est calculer sur des symboles comme dans un système formel. Dès lors, une "science de l'esprit" se doit d'être formelle au sens que l'on donne à ce terme en logique.

On comprend de ce fait que, hors du cadre calculatoire qui définit la méthode de l'intelligence artificielle, il n'y a pas de sens précis aux notions qui ne seraient pas présentées selon ce format. C'est la raison pour laquelle, jusqu'à présent, aucune objection d'inspiration non-logique (en particulier,

---

[53] Cf. **G. Frege**, *Begriffschrift*, traduction anglaise dans [**van Heijenoort J. ed.** (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Cambridge], § 13 : «Nous avons déjà introduit un certain nombre de principes fondamentaux de pensée dans le premier chapitre pour les transformer en règles en vue de l'usage de nos signes. Ces règles et les lois dont elles sont les transformations ne peuvent pas être exprimées dans l'idéographie parce qu'elles forment leur fondement».

[54] Cf. **Hobbes**, *Leviathan*, première partie, chap. V. Le texte exact dit d'ailleurs tout autre chose : «La Raison n'est rien d'autre que l'estimation [*reckoning*]».

phénoménologique) adressée à l'intelligence artificielle n'a eu d'incidence réelle sur la conception même du projet.

12. La sémantique empiriste du logicisme russellien

La forme primitive de la notion de système formel a été modifiée à partir des difficultés sémantiques liées aux paradoxes de la théorie naïve des ensembles. Les paradoxes, en particulier celui de Russell[55] touchant l'ensemble de tous les ensembles ne se contenant pas eux-mêmes comme éléments, exigent que soit repensée la définition de la notion d'ensemble : cette question contraint à envisager autrement l'un des traits spécifiques de la notion de système formel, celle de l'universalité de son domaine d'application. En effet, pour réussir à éliminer les paradoxes, il faut éliminer la possibilité de construire au sein du système formel des ensembles contradictoires. C'est ce but que se propose de réaliser la théorie des types de Russell. Pour préserver l'universalité d'application du système formel tout en éliminant les risques de contradiction, il faut introduire des types, c'est-à-dire des différences entre classes d'objets, au sein desquelles tous les objets sont traités de la même manière. Ainsi l'universalité de la logique est-elle préservée “par palier”.

La difficulté de cette théorie provient de ce qu'elle bute sur des notions irréductibles à des considérations sur les types. Comme le font remarquer Kneale et Kneale[56], c'est le cas des théorèmes portant sur les nombres réels en général. Si l'on définit un nombre réel à la manière de Dedekind comme formé par une coupure sur les rationnels, c'est-à-dire comme un ensemble infini de nombres rationnels, on se trouve confronté à la difficulté suivante. Si S est un ensemble non-vide de nombres réels, son plus petit majorant est un nombre réel qui a pour membre tous les nombres rationnels qui sont membres de n'importe quel membre de S. Dans ce cas, l'expression fonctionnelle qui spécifie cette borne supérieure doit faire référence à la totalité des fonctions propositionnelles qui spécifient des nombres réels, y compris la fonction qu'elle exprime. On ne peut donc pas

---

[55] Lettre de Russell à Frege (1902), traduction française dans [**F. Rivenc** et **P. de Rouilhan dirs.**, *Logique et Fondements des mathématiques*, Anthologie (1850-1914), op. cit.], pp. 240-241.

[56] **W. Kneale** et **M. Kneale**, *The Development of Logic*, Clarendon Press, Oxford, 1962, p. 662.

spécifier de type. Dès lors, seule l'introduction de l'axiome de "réductibilité" permet à Russell de justifier l'existence de théorèmes de l'Analyse portant sur les nombres réels en général. Celui-ci apparaît alors comme un axiome *ad hoc*. Les considérations sémantiques reçoivent ainsi un statut seulement empirique[57] qui ne sont justifiées que par l'idée provisoire selon laquelle on pourra toujours modifier le système formel si la nécessité s'en fait sentir. Le système formel a donc le statut d'une généralisation inductive que l'on espère être complète, sans jamais en être certain.

C'est la signification russellienne de la notion de sémantique au sein d'un système formel qui est reprise à son compte par l'intelligence artificielle. En effet, dans la comparaison entre le cerveau humain et un système formel, on considère aussi que le sens des symboles du système formel est donné empiriquement.

Puisqu'un système formel est autonome par rapport à toute détermination extérieure (c'est-à-dire que son fonctionnement n'est régi que par ses règles propres), on doit considérer l'esprit comme un objet autonome ne recevant pas de l'extérieur ce qui permet d'assurer son contrôle. Dès lors, la notion de sémantique qui, dans un système formel, est radicalement distinguée de celle de syntaxe, est-elle aussi distinguée dans le cas de l'esprit : le rapport à la réalité extérieure devient ainsi une donnée purement empirique qui n'influe en rien sur le fonctionnement de l'esprit lui-même[58]. On sait que si une interprétation d'un système formel rend les axiomes vrais et engendre des propositions valides selon les règles d'inférence du système, le système formel préservera toujours la vérité des propositions. C'est pourquoi en intelligence artificielle, on se permet de séparer radicalement les questions syntaxiques des questions sémantiques, puisque le système formel se charge lui-même d'opérer des inférences valides qui sont

---

[57] **A. Lautman**, "Considérations sur la logique mathématique" dans [*Essai sur l'unité des mathématiques et divers écrits*, Union Générale d'Édition, Paris, 1977], pp. 308-309 : «C'est par cet appel constant à l'expérience dans le choix de notions premières et dans l'introduction des différentes opérations logiques que Russell et Whitehead sont certains d'avoir éliminé à l'avance les contradictions et les paradoxes».

[58] Ce point de vue est décrit dans **J. Haugeland**, "Semantic Engines : an Introduction to Mind Design", dans [**J. Haugeland ed.**, *Mind Design*, MIT Press, Cambridge, 1981], pp. 9-10.

vraies si les axiomes ont été correctement choisis[59]. Bref, une “bonne” sémantique dépend d’une “bonne” syntaxe et la première découle toujours de la seconde.

La question qui demeure est bien entendu celle de la vérité des axiomes. Mais l’établissement de la vérité des axiomes reste, dans cette optique, une question empirique. C’est en effet par essais et erreurs que l’on tente de conférer une interprétation à un système formel et à la totalité de ces théorèmes : si l’interprétation n’est pas recevable parce qu’elle paraît “à l’évidence” fausse, on doit changer d’interprétation. Cependant le critère formel de cette certitude est laissé dans l’ombre et ne relève que de considérations purement empiriques : si l’interprétation paraît adéquate, on la garde jusqu’à nouvel ordre. C’est pourquoi l’intelligence artificielle apparaît comme une forme particulièrement abstraite de psychologie[60]: ce qui fait sens peut être modélisé par le biais de la programmation sur ordinateur et toutes les expériences de psychologie, si sujettes à caution, peuvent être éliminées.

On remarque cependant que les notions centrales de l’intelligence artificielle dans sa version calculatoire et symbolique, comme celle de calcul ou de représentation, ne sont pas héritées du cadre de pensée logiciste alors qu’elles constituent l’armature théorique du projet d’intelligence artificielle[61]. C’est le formalisme hilbertien qui a forgé ces notions pour répondre à ses besoins propres.

2. Ce que l’intelligence artificielle doit au formalisme

---

59 Cf. **D. Dennett**, “Three Kinds of Intentional Psychology” dans [R. A. Healey, *Reduction, Time and Reality : Studies in the Philosophy of natural sciences*, Cambridge, Cambridge University Press, 1981].

60 Cf. C’est ce que fait remarquer Haugeland dans **J. Haugeland**, “Semantic Engines : an Introduction to Mind Design”, dans [**J. Haugeland ed.**, *Mind Design*, MIT Press, Cambridge, 1981], pp. 31.

61 Le moins que l’on puisse dire, c’est que l’intérêt de la philosophie de l’esprit de tendance analytique pour l’œuvre de Frege est sélectif : Frege niait la possibilité même d’une psychologie des mathématiques puisque l’objectivité des concepts mathématiques ne pouvait dépendre, pour lui, de la subjectivité des individus. Cf. par exemple, **G. Frege**, “Recherches logiques. 1. la Pensée”, traduction française dans [**G. Frege**, *Écrits logiques et philosophiques*, Le Seuil, Paris, 1971], pp. 170-171.

La notion de calcul telle qu'elle a été élaborée par Turing et son application première à un problème mathématique, le "problème de la décision"[62], sont nées au sein d'un cadre épistémologique formaliste. C'est en effet pour répondre à ce problème posé par Hilbert que Turing élabora son concept original de machine. Le problème de la décision est le suivant : existe-t-il une procédure systématique et effective qui permettrait de déterminer à propos d'un énoncé mathématique parfaitement spécifié s'il est ou non démontrable ? A cette question, Turing répond par la négative. Turing résout donc le problème de la décision tel qu'il apparaît dans un cadre hilbertien. Dance ce cadre, l'interprétation des mathématiques telle qu'elle est contenue dans ce qu'il est convenu d'appeler "le programme de Hilbert"[63], est très différente de l'approche logiciste.

En quoi consiste la formalisation ? Comme le fait remarquer J. Largeault, elle ne consiste pas en une réduction des mathématiques à la logique [64]:

> **«Elle consiste à reconstruire simultanément logique et mathématiques, en maintenant l'autonomie des deux disciplines pour une partie élémentaire de l'arithmétique, dont les concepts et les inférences intuitives devraient suffire à prouver la non-contradiction formelle de tout le reste, uniformément transcrit en langage symbolique».**

Se posent alors dans le cadre formaliste des types de problèmes qui lui sont spécifiques : les notions de complétude et de consistance d'un système formel prennent un sens qu'elles ne pouvaient avoir dans le cadre du logicisme, parce que la notion de système formel peut être envisagée de l'extérieur, d'un point de vue méta-systématique, alors que l'universalisme de la logique telle qu'il était conçu

---

[62] Il est exposé par Hilbert dans "Problèmes de fondation des mathématiques", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, Vrin, Paris, 1992], p. 183. Pour des remarques historiques sur les solutions partielles apportées au problème avant les démonstrations indépendantes de Church et de Turing, cf. **J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit., p. 176-177.

[63] J. Largeault résume ainsi, dans un vocabulaire moderne, le but du "programme de Hilbert" : «En simplifiant, le programme de Hilbert repose sur l'idée que les théorèmes mathématiques sont mécaniquement prouvables et que vérifier la correction des preuves est une tâche d'ordinateur». Présentation de la traduction française de l'article de Hilbert "Les fondements des mathématiques", dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 145.

[64] **J. Largeault éd.**, *Intuition et Intuitionisme*, Vrin, Paris, 1993, p. 75.

par le logicisme interdisait de considérer autre chose que l'univers du discours dans son entier[65].

D'un point de vue formaliste, on doit être capable, à partir d'un système fini de signes définis à l'avance et manipulés selon des règles abstraites, d'écrire tout énoncé mathématique en le retranscrivant symboliquement et de rendre compte de toutes les opérations mathématiques qu'on est capable d'exécuter sur lui. Ainsi peut-on, pour Hilbert, fournir une image des mathématiques intuitives, image constituée par un jeu logique entre des formules. Hilbert déclare [66]:

> **«L'idée maîtresse de ma théorie est la suivante : tout ce qui constitue les mathématiques au sens traditionnel est rigoureusement formalisé, en sorte que la mathématique proprement dite ou au sens strict devient un stock de formules. Celles-ci se distinguent des formules qu'on a l'habitude d'utiliser en mathématiques seulement par le fait qu'en dehors des signes usuels, y apparaissent en plus des signes logiques, en particulier ceux pour "implique" (->) et "ne … pas" (¬)».**

Une telle attitude a des conséquences fondamentales sur l'interprétation à accorder aux notion de calcul, de représentation et plus généralement, à ce que l'on doit entendre par l'expression d'opération de l'esprit.

Commençons par l'idée de représentation.

21. La notion de représentation dans l'épistémologie formaliste

La distinction de l'aspect sémantique et syntaxique des énoncés implique de posséder, au sein du système formel, une représentation des énoncés mathématiques[67]. La formalisation a précisément pour but de représenter des énoncés mathématiques ayant intuitivement un contenu sémantique par des marques symboliques dénuées de toute interprétation et pour lesquelles seul l'agencement syntactique sera pris en compte. Dès lors, ces marques symboliques jouent le rôle de représentation pour les énoncés mathématiques "intuitifs", c'est-

---

65 Cf. **J. Van Heijenoort**, "Logic as calculus and Logic as language" dans [*Selected Articles*, Bibliopolis, Napoli, 1985], pp. 13-14.

66 **D. Hilbert**, "Le fondement de l'arithmétique élémentaire", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 191-192.

67 C'est ainsi que Kreisel décrit le sens général du "programme de Hilbert" : «Le programme de Hilbert au sens large se proposait d'établir "l'adéquation" de formalismes déductifs à la représentation de branches intuitives des mathématiques». **G. Kreisel**, "Le programme de Hilbert", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 485-486.

à-dire pour les énoncés au sein desquels on n'a pas encore opéré de séparation entre l'aspect sémantique et l'aspect purement symbolique.

L'intuition est dès lors limitée à sa fonction cognitive et n'apparaît plus comme une faculté métaphysique : l'intuition est la faculté psychologique qui permet de distinguer des marques symboliques écrites sur un support matériel *et rien de plus*. Hilbert fait remarquer à ce sujet[68]:

> **«La condition préalable de l'application des inférences logiques et de l'effectuation d'opérations logiques est l'existence d'un donné de la perception : à savoir l'existence de certains objets concrets extra-logiques qui en tant que sensations immédiates précèdent toute pensée».**

La théorie mathématique de la représentation apparaît alors comme une théorie du codage formel des énoncés mathématiques[69]. Cette idée de codage formel des énoncés sera poursuivie d'une part par Gödel qui, à chaque énoncé mathématique formalisé associe un nombre entier sur lequel il est possible de faire porter un calcul et d'autre part par Turing qui se sert explicitement de l'idée du codage numérique gödelien dans son exposition de la solution du problème de la décision par le biais de son concept de machine.

Ainsi l'introduction de la notion de représentation dans le débat sur la

---

68 **D. Hilbert**, "Sur l'infini", traduction française dans [**J. Largeault éd.**, *Logique mathématique,* textes, Armand Colin, Paris, 1972] p. 228. Cependant, ce finitisme "positiviste" possède une contrepartie "métaphysique" dans la mesure où Hilbert est bien forcé de reconnaître que l'attitude finitiste est suffisante pour décrire les quatre opérations arithmétiques primaires, qui ont bien le statut de concepts virtuels et non pas seulement de signes actuels. Cf. **D. Hilbert**, "Nouvelle fondation mathématique. Première communication", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 117-118. Il faut remarquer que ces «sensations immédiates qui précèdent toute pensée» son aussi appelées "représentations" par Hilbert, dans le sens que ce terme possède dans la tradition philosophique et non dans le sens technique qu'il lui donne en mathématique; il dit par exemple : «(…) à titre de condition préalable pour l'application d'inférences logiques, quelque chose nous est déjà donné dans la représentation : certains objets extra-logiques qui, en tant qu'expérience immédiate sont intuitivement présents avant toute pensée». **D. Hilbert**, "Les fondements des mathématiques", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 147.

69 Comme le remarque J. Largeault : «Qu'il fût possible de représenter les raisonnements et conceptualisations mathématiques par un système de règles mécaniques capable d'engendrer les répliques formelles des théorèmes et de leurs démonstrations a été surprise et une grande découverte». Présentation de la traduction française de l'article de Hilbert "Nouvelle fondation mathématique. Première communication", dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 108. On trouve une analyse formelle de la notion de représentation dans **R. M. Smullyan**, *Theory of Formal Systems*, Annals of Mathematics Studies n° 47, Princeton University Press, 1961, chap. 1, § 3.

nature des mathématiques est-elle intimement liée à la constitution du programme formaliste et de son axiomatisation[70], contrairement à ce que l'on aurait pu être tenté de croire si l'on en était resté à une analyse philosophique en termes logicistes. Aussi l'accent mis sur les représentations symboliques en intelligence artificielle[71] est-il directement hérité de la conception hilbertienne des mathématiques.

22. La notion de calcul dans l'épistémologie formaliste

C'est par rapport à la question du finitisme en mathématique que s'élabore progressivement au sein du formalisme une recherche précise sur ce qu'il faut entendre par calcul, alors que cette réflexion n'avait pas été perçue comme nécessaire dans le cadre logiciste. Là encore, la dette de l'intelligence artificielle à l'égard de la conception hilbertienne des mathématiques est grande : la théorie calculatoire de la manipulation des symboles dérive entièrement du point de vue finitiste.

Pour Hilbert, le contenu sémantique des énoncés mathématiques, c'est-à-dire leur vérité, est éliminé du formalisme et représenté par leur non-contradiction logique. Plus précisément, il faut essayer de montrer que la mathématique classique dans son entier, une fois axiomatisée, peut s'interpréter dans le formalisme logique particulier de la logique des prédicats du premier ordre. Il s'agit là d'un projet qu'il faut tenter de vérifier; aussi le projet doit-il être considéré comme une thèse, baptisée depuis dans la littérature "thèse de Hilbert"[72]: les axiomes d'une théorie mathématique quelconque peuvent être exprimés dans le cadre de la logique du premier ordre et en particulier, la notion

---

[70] J. Largeault fait remarquer à ce sujet : «Les formules d'une théorie axiomatique peuvent être codées en nombres entiers naturels au moyen d'une numérotation. Par ce détour, les notions de récursivité et de T[uring]-calculabilité s'appliquent à des objets de nature non-numérique. Une théorie est axiomatisable si l'ensemble (des nombres de Gödel) de ses théorèmes est récursivement énumérable». **J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit., p. 243.

[71] Voir par exemple sur l'importance accordée aux représentations en intelligence artificielle, **M. Denis**, "Pour les représentations", dans [**M. Denis** et **G. Sabah dirs.**, *Modèles et Concepts pour la science cognitive*, hommage à J. -F. Le Ny, Presses Universitaires de Grenoble, Grenoble, 1993], pp. 95-106. L'article ne rapporte cependant pas l'introduction du concept à son origine formaliste.

[72] Cette appellation est due à Martin Davis. Cf. **J. Barwise ed.**, *Handbook of Mathematical Logic*, North-Holland, Amsterdam, 1977, p. 41.

informelle de “démontrable” devient précise grâce à la notion de “démontrable en logique du premier ordre”. L’accent mis sur la démontrabilité vient de ce que, toute intuition ayant été laissée de côté, une proposition mathématique n’est recevable, une fois retranscrite comme proposition du système formel, que si on peut la dériver des axiomes du système. Comment savoir si une formule est ou non dérivable à partir des axiomes du système formel ? Il faut pouvoir s’en assurer, au moins en droit, en possédant un moyen de contrôle permettant de parvenir à établir la légitimité de cette dérivation.

La difficulté qui se présente consiste alors à éliminer les expressions faisant appel à des domaines de valeurs infinis parce qu’on ne peut pas les passer en revue de façon finie. En particulier, dès qu’une expression possède des quanteurs, elle ne peut recevoir de statut formel que s’il y a un moyen d’éliminer les quanteurs et de les évaluer effectivement[73]. C’est l’exigence d’effectivité au sein du formalisme qui va conduire les mathématiciens formalistes à rechercher une détermination précise de ce que l’on entend par calcul et à mettre au jour la notion d’algorithme (ou de fonction calculable), comme nous le verrons plus loin. L’apport de Turing consiste précisément à avoir montré qu’il était possible d’assimiler une inférence à un calcul : le contrôle de la dérivation au sein d’un système formel consiste donc à n’admettre au rang de proposition dérivable que celle dont on peut justifier l’engendrement par le biais d’un algorithme de calcul. La notion de calcul se situe donc au cœur de la notion de système formel dans la mesure où c’est sur elle que repose, en dernière instance, l’idée de l’existence d’une méthode logique de preuve du vrai[74].

C’est en particulier le statut des propositions infinitaires de l’Analyse qui

---

[73] J. Largeault donne l’exemple suivant : «Pour tout entier x il existe un entier y tel que A(x, y) est vrai, a un sens finitiste quand il existe une fonction calculable y = f(x) et un moyen de décider, pour chaque couple (x, y), si y entretient avec x la relation A». Présentation de la traduction française de l’article de Hilbert “Nouvelle fondation mathématique. Première communication”, dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 109.

[74] Il y a un rapport d’analogie entre la notion d’algorithme et celle de système formel. Comme le fait remarquer J. Largeault : «N’importe quel algorithme générateur de suites de signes peut compter pour un analogue formel d’une opération productrice de conséquences; si ces suites ne sont susceptibles d’aucune interprétation conceptuelle ou propositionnelle, cela ne vaudra pas pour une logique.» **J. Largeault**, *La logique*, Presses Universitaires de France, Paris, 1993, p. 55.

fait question. Hilbert fait remarquer à ce propos[75] :

> **«Dans le cas d'une infinité d'objets, la négation de l'énoncé général (a) A(a) n'a, pour commencer, aucun contenu précis, pas plus que la négation de l'énoncé d'existence (Ea) A (a). […]. Pour les collections finies, "il y a" et "est donné" signifient la même chose; pour les collections infinies, seul le second de ces deux concepts est directement clair. Nous voyons donc que si l'on se propose de fonder rigoureusement les mathématiques, on ne peut pas accepter comme allant de soi au point de vue logique les formes d'inférence couramment utilisées en Analyse. […] En restant sur le terrain finitiste, il s'agit donc d'arriver à manier librement et à dominer entièrement le transfini.»**

Trois concepts doivent donc entretenir des rapports cohérents : le calcul, le continu et le formel. Cette cohérence doit être atteinte par une théorie adéquate de la représentation des énoncés mathématiques, en particulier ceux qui portent sur le transfini. C'est en ce sens que le problème de la maîtrise du continu revêt une importance capitale et que Hilbert y voyait la "pierre de touche" de son programme formaliste[76].

La polémique qui s'instaura entre l'école formaliste de Hilbert et l'école intuitioniste de Brouwer a ainsi pour objet la nature de la représentation des énoncés mathématiques[77].

23. Représentation du continu et représentation de la pensée dans l'épistémologie formaliste

Une première façon d'interpréter mathématiquement la notion de continu consiste

---

75 **D. Hilbert**, "Les fondements logiques des mathématiques", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 136.

76 Cf. **D. Hilbert,** "Sur l'infini", traduction française dans J. **Largeault éd.**, *Logique mathématique,* textes, op. cit., p. 237 : «Je voudrais encore faire allusion à une dernière victoire. Pour toute théorie nouvelle, la pierre de touche est le succès dans la résolution de questions qui préexistent à cette théorie en vue desquelles elle n'avait pas été créée. "Vous les reconnaîtrez à leur fruit", cela s'applique aussi aux théories. Aussitôt que Cantor eut découvert ses premiers nombres transfinis, ceux qu'on appelle nombres de la seconde classe, s'est posée la question, que j'ai mentionnée plus haut, de savoir si cette énumération transfinie permettrait aussi d'énumérer des ensembles connus par ailleurs et qui , au sens courant du mot, ne sont pas dénombrables. Comme exemple d'un tel ensemble, l'intervalle de points se présenta en premier».

77 Comme le fait remarquer J. Largeault, c'est plus la réflexion sur la nature du continu et du discret qui définit l'intuitionisme de Brouwer que la polémique avec Hilbert sur le rôle du tiers-exclu. Cf. **J. Largeault**, *L'intuitionisme*, Presses Universitaires de France, Paris, 1992, p. 21 : «Le constructivisme, en dépit de ce que Hilbert croyait ou feignait de croire (…) n'est peut-être pas un trait essentiel de la doctrine de Brouwer, qui se caractérise plutôt par la manière de réagir à l'aporie du discret et du continu, l'admission d'êtres partiellement indéterminés, le renoncement aux facteurs de détermination verbaux de la logique classique, enfin par ses bases métaphysiques».

à considérer le continu comme un ensemble de points, selon les enseignements de la théorie cantorienne des ensembles. C'est cette théorie que veut conserver et défendre le programme formaliste en proposant une solution au problème du continu de Cantor[78].

Une fois admis que le continu est composé d'un ensemble de points, on se rend compte que la totalité de ces points est indénombrable et que le calcul ne peut en rendre accessible qu'une partie - une infinité dénombrable - et ne peut dès lors constituer qu'un "continu réduit"[79], celui des réels calculables. Cette conception suppose donc l'existence d'un classe numérique partiellement inaccessible, celle des nombres réels en général.

La question de la représentation du continu dans le cadre formaliste, qui revient au problème de la représentation logique des énoncés mathématiques portant sur le transfini, ressort bien dès lors du problème de la représentation en général, tel qu'il nous est déjà apparu[80] : en effet, de même que l'idée d'une représentation des phénomènes internes et externes relève d'un mouvement qui consiste à délimiter une partie de la nature tout en y étant soi-même plongé, de même ici, ce qu'il est possible de se représenter du continu correspond au continu réduit, c'est-à-dire à une délimitation du continu qui se trouve également plongé dans le continu entier. Étudions du point de vue de la théorie de la représentation ce passage du continu au continu réduit. Du point de vue de l'aspect objectif de la représentation, le continu acquiert un statut représentationnel sous une forme "réduite"; du point de vue de l'aspect subjectif de la représentation, il devient possible de rendre accessible à la pensée l'acte de pensée propre au calcul. Ainsi ce qui se manifeste dans la représentation ne peut-il

---

[78] Pour Cantor, il s'agissait au départ de savoir quelle est la cardinalité de l'ensemble formé par les points d'un intervalle de la droite réelle. Cette question, somme toute particulière, revêt cependant une grande importance dans la mesure où c'est elle qui permettrait de décider s'il est légitime de concevoir une échelle des cardinalités, maîtrisable par une arithmétique appropriée et dont l'opération fondamentale serait celle de mise à la puissance. Réussir à déterminer la taille du continu apparaît alors comme une stratégie indirecte en vue de faire avancer la réponse à la question de la détermination généralisée de la taille des ensembles. Cf. **M. Hallet**, *Cantorian set theory and limitation of size*, Clarendon Press, Oxford, 1984, p. 86 sq.

[79] L'expression de "continu réduit" est de Brouwer. Cf. **J. Largeault**, *Intuition et Intuitionisme*, op. cit., p. 48.

[80] Cf. supra, Avant-propos, p. 6.

apparaître qu'à l'intérieur de deux limites, l'une objective et l'autre subjective : la représentation ne manifeste d'un point de vue objectif que le continu réduit et d'un point de vue subjectif que l'aspect finitiste des actes de pensée. On comprend dès lors que Hilbert interprète la notion de représentation en mathématique comme ce qui relève de l'effectivité finitiste telle qu'elle se déploie dans l'arithmétique des nombres entiers *mais aussi que la notion de représentation en général soit pour lui de même nature*. C'est pourquoi Hilbert écrit [81]:

> **«[…] c'est que notre pensée est finitiste; quand nous pensons, se déroule un processus finitiste».**

Ainsi passe-t-on de la notion de représentation des mathématiques à la notion de représentation de la pensée quand on se heurte au problème de la maîtrise mathématique du continu[82].

De même, Turing hérite directement de ce passage du mathématique au psychologique *via* la notion de représentation. La notion de calcul dans un cadre formaliste se trouve définie chez lui à partir de la question de la calculabilité des réels, c'est-à-dire la question de la maîtrise arithmétique du continu : le concept de machine tel qu'il est inventé par Turing se situe ainsi à l'articulation du discret des opérations de calcul et du continu des nombres réels et c'est pourquoi ce concept opère le passage entre représentation mathématique du continu et représentation de la pensée.

On voit combien le projet même de l'intelligence artificielle est lié au cadre de pensée hilbertien puisque c'est d'abord en lui que le modèle discret d'une représentation de l'esprit est tout d'abord apparu. Il devient en particulier possible, au sein du formalisme, de circonscrire le domaine de la pensée par le biais du concept de représentation et partant, de rompre avec l'adhérence de la pensée à

---

[81] **D. Hilbert**, "Les fondements des mathématiques", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 140.

[82] Le passage du domaine proprement mathématique au domaine philosophique se note aussi dans la capacité d'invention liée à la maîtrise du continu. Comme le remarque Wang Hao, c'est en cherchant une preuve de cohérence de l'analyse que Gödel a trouvé son résultat d'incomplétude de l'arithmétique; cf **W. Hao**, "Some Facts about Kurt Gödel", Journal of Symbolic Logic, XLVI-1981, p. 654.

elle-même, principal obstacle à la constitution d'une théorie de la psychologie[83].

3. Ce que l'intelligence artificielle doit à l'intuitionisme

L'intelligence artificielle semble ne pas avoir de dette directe à l'égard de l'intuitionisme de Brouwer, selon lequel la pratique des mathématiques relève d'un acte non-linguistique et non pas de l'étude des suites réglées d'opérations sur des signes[84]. Cette conception des mathématiques engendre en effet une critique radicale de la théorie de la représentation telle qu'elle se constitue au sein du formalisme. Brouwer fait remarquer à ce sujet [85]:

> **«[…] entre la perfection du langage mathématique et la perfection des mathématiques proprement dites, on ne peut discerner aucune relation évidente».**

Dans ces conditions, le projet même du formalisme et son lointain descendant, le projet de l'intelligence artificielle, semblent compromis puisque la notion de calcul formel sur des représentations fait complètement défaut.

Remarquons cependant que la notion de calcul *ne varie pas* quand on passe de la perspective formaliste à la perspective intuitioniste. Comme le fait remarquer J. Largeault, on peut se demander alors si la querelle entre intuitionistes et formalistes n'est pas vaine quand on en vient à la détermination de la nature du

---

[83] Cf. supra, Première partie, Introduction, § 1. Essayer, comme c'est le cas pour l'instant dans la réflexion philosophique sur le statut scientifique de l'intelligence artificielle, de ne conserver du formalisme de la machine de Turing que son aspect discret en occultant la finalité du formalisme qui vise à circonscrire la classe des réels calculables, et partant le continu, me paraît donc avoir des conséquences dommageables sur le débat concernant la nature de la notion de représentation en intelligence artificielle.

[84] C'est ce que Brouwer appelle le "premier acte de l'intuitionisme" : «Pour commencer, le premier acte de l'intuitionisme sépare complètement les mathématiques du langage mathématique, en particulier des phénomènes de langage que décrit la logique théorique, et reconnaît que la mathématique intuitioniste est une activité de l'esprit essentiellement sans langage, qui prend son origine dans la perception d'un coup de temps, *i. e.* de la division d'un instant de vie en deux objets distincts, dont l'un donne naissance à l'autre tout en étant conservé par le souvenir». **L. E. J. Brouwer**, "Base historique, principes et méthodes de l'intuitionisme" traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 449.

[85] **L. E. J. Brouwer**, "Base historique, principes et méthodes de l'intuitionisme" traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 448. J. Largeault fait remarquer que «Brouwer s'est abstenu d'employer les notations canoniques introduites par les logiciens (…). il pensait que le langage ordinaire est plus évocateur parce que plus chargé d'images et plus propre à susciter la réflexion». **J. Largeault**, *Intuition et intuitionisme*, op. cit., p. 117 note 1.

concept de calcul[86]:

> **«Les précisions apportées par Gödel, Church, Kleene et Turing, sur la calculabilité pouvaient sembler rendre inutile le constructivisme intuitioniste, puisqu'une sorte d'équivalent mécanique de la constructivité - la calculabilité ou la récursivité - se laisse définir dans le cadre des notions et principes logiques classiques».**

Dans ces conditions, ce que montrerait l'intuitionisme, du point de vue de l'intelligence artificielle, c'est qu'il est possible de changer d'ontologie sans changer de concept de calcul et qu'il est donc possible d'*affranchir le concept de calcul de toute thèse ontologique* touchant le mode d'être de l'objectivité mathématique.

De ce point de vue, le rapprochement que nous avons fait entre l'intelligence artificielle et la perspective formaliste serait à la fois vrai historiquement et sans importance du point de vue de l'intelligence artificielle en tant que projet scientifique positif : l'intelligence artificielle aurait hérité d'un certain nombre de concepts qui sont apparus au sein de la perspective formaliste mais ceux-ci ne lui seraient pas intrinsèquement liés. Le concept de calcul serait l'un d'eux, dans la mesure où il ne serait ni formaliste ni intuitioniste : il y aurait donc bien une dette indirecte de l'intelligence artificielle à l'égard de l'intuitionisme, qui proviendrait moins d'un héritage conceptuel proprement dit que d'une alternance philosophique à la perspective formaliste, alternance qui renforcerait l'idée qu'il est possible d'isoler la notion de calcul de toute prise de position ontologique. Dès lors, le souci philosophique du projet d'intelligence artificielle touchant la notion de calcul serait d'en faire un outil épistémologique libéré de toute "entrave" ontologique qui en contrarierait la positivité.

C'est précisément ce point de vue philosophique "positif" qu'il faut tenter d'examiner. Car l'intelligence artificielle peut-elle vraiment se passer de toute perspective ontologique ? Et le concept de calcul n'est-il qu'un concept opératoire qui ne véhicule pas de lui-même une certaine ontologie ? Il faut, pour répondre à ces questions, revenir un instant sur l'alternance ontologique que propose l'intuitionisme de Brouwer à la perspective formaliste. On remarque en effet que

---

[86] **J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit., p. 422.

ce n’est pas sur la nature du concept de calcul mais sur le rapport du calcul au continu que porte la différence entre l’ontologie formaliste et l’ontologie intuitioniste. C’est donc sur ce terrain qu’il faudra par la suite examiner si le projet de l’intelligence artificielle peut effectivement se passer d’ontologie.

31. Calcul et continu dans l’intuitionisme de Brouwer

C’est la question de la nature du concept de fonction qui distingue les deux perspectives. Comme le fait remarquer J.-M. Salanskis, la perspective formaliste offre la possibilité d’envisager les fonctions sous l’aspect de leur discontinuité, ce qui n’est pas autorisé d’un point de vue intuitioniste[87].

En effet, dans le cadre ensembliste qui est celui de la perspective formaliste[88], les fonctions continues sont l’exception : à chaque fonction numérique au voisinage d’un point correspond une infinité de graphes fonctionnels parmi lesquels une infime minorité sont les graphes de fonctions continues. Or c’est précisément sur ce point que se distingue l’intuitionisme de Brouwer puisqu’une fonction bien définie y est nécessairement considérée comme uniformément continue. Le concept de calcul et la notion de continu entretiennent ainsi de nouvelles relations réciproques qui ne sont pas médiatisées par la notion d’ensemble. Examinons ces relations.

Si le continu était un ensemble de points, il ne serait jamais accessible en tant que tel, sauf sous la forme du continu réduit[89]. Mais le continu réduit n’est, pour Brouwer, que le dépôt linguistique du continu au sens plein.

Le calcul entretient avec le continu des relations de réciprocité, dès lors que l’on conçoit ce dernier non pas comme un ensemble mais comme le résultat d’un acte effectif de pensée aux prises avec un objet mouvant. Comme le

---

[87] **J.-M. Salanskis**, “Le destin du modèle de Cantor-Dedekind” dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, Springer-Verlag, Paris], pp. 203-204 : «Il est clair que la grande affaire du constructivisme est la notion de *fonction* : c’est en elle que se joue à la fois la supériorité et l’infériorité de l’approche constructive.».

[88] **J.-M. Salanskis**, *L’herméneutique formelle*, Presses du CNRS, Paris, 1991, pp. 151-153. L’auteur cite le cas de Lagrange qui introduit l’analycité comme caractéristique de toute fonction.

[89] Brouwer le dit explicitement : «Le continu en tant qu’ensemble de points n’existe pas». **L. E. J. Brouwer**, “Over de Grondslagen der Wiskunde”, 1907, p. 83, cité dans [**J. Largeault éd.**, *Intuition et intuitionisme*, op. cit.], p. 147.

remarque J. Largeault [90]:

> **«Si la solution [d'un problème bien posé] dépend continûment de paramètres qui expriment les conditions initiales, elle doit être calculable à partir des valeurs données à ces paramètres. Ainsi la continuité impliquerait la calculabilité (la constructivité).»**

Réciproquement, le calcul produit le continu. Si l'on prend le cas du calcul des nombres réels grâce au concept de suite, définie comme engendrement successif des nombres qui la comprenne, on distingue, dans le concept de suite, les suites déterminées jusqu'à l'infini par une loi et les suites libres de choix dont on ne peut décider à l'avance si elles possèdent telle ou telle propriété (par exemple, si la suite, composée d'une infinité de nombres, possède ou non le nombre 1). Dès lors, la notion de suite déterminée représente - mais dans un sens non formaliste - un nombre réel particulier tandis que la notion de suite de choix libre représente le continu. C'est ce que fait remarquer H. Weyl [91]:

> **«C'est une première idée fondamentale de Brouwer que la suite des nombres qui croît par des actes libres de choix est un objet possible de conceptualisation mathématique. Si la loi qui définit une suite jusque dans l'infini représente le nombre réel particulier, de même la suite de choix, dont aucune loi ne restreint la liberté de développement, représentera le continu. […] Plus généralement peut s'utiliser à cette même fin toute loi suivant laquelle, dans une suite en devenir de nombres naturels, chaque choix qui y ajoute un terme engendre un nombre déterminé. Ainsi le nombre engendré au h-ième pas ne dépendra pas seulement du choix effectué au h-ième pas, mais de tout le segment initial de la suite de choix arrêtée à cet instant, et allant de son premier à son h-ième terme. […]. La remarque de Brouwer est simple, mais profonde : là apparaît un "continu" dans lequel tombent, certes, les nombres réels isolés sans qu'il se résolve en un ensemble de nombres réels dont l'existence serait achevée; bien plutôt s'agit-il d'un *milieu de libre devenir*».**

Un continu émerge du fait que l'on ne rapporte pas son engendrement à la *règle* de construction qui préside à la constitution de la suite mais à un *intervalle* de points dont le caractère lié dépend d'une suite de choix. On comprend alors que l'effectif puisse se concilier avec l'indéterminé : des objets mathématiques peuvent exiger une infinité potentielle pour faire émerger leurs propriétés;

---

[90] **Largeault J.**, *Intuition et intuitionisme*, op. cit.], p. 120.

[91] **H. Weyl**, "Sur la crise contemporaine des fondements des mathématiques", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 76-77. Cf. aussi J. Largeault : «(…) si pour une suite de choix α on peut calculer un nombre naturel $F(\alpha, b)$, alors si β est une suite de choix qui coïncide avec α sur un segment initial fini, F doit associer à β le même nombre naturel b, autrement dit F est continue». **J. Largeault éd.**, *Intuition et intuitionisme*, op. cit., p. 123.

autrement dit, des objets mathématiques sont susceptibles d'acquérir de nouvelles propriétés.

On voit donc que le même concept de calcul peut entraîner comme corrélat ontologique des univers radicalement différents - définis par leur aspect continu ou discontinu - dès que l'on réfère la notion de fonction à celle de procédé de calcul ou à celle d'ensemble. Ces différences ontologiques ont des conséquences sur la façon de concevoir la nature de la pensée d'un point de vue intuitioniste.

32. Représentation du continu et représentation de la pensée dans l'épistémologie intuitioniste

Une idée demeure chez Brouwer que l'on avait déjà rencontrée chez Hilbert : il y a un rapport étroit entre la conception mathématique que l'on se fait du continu et l'idée que l'on se fait de la démarche de la pensée. Mais ce rapport n'est plus médiatisé chez Brouwer par la notion de représentation formelle mais par l'intuition originaire d'un acte temporel non-linguistique [92]:

> **«Le néo-intuitionisme considère la dissociation d'instants vécus en parties qualitativement distinctes, qui ne se réunissent qu'en restant séparées par le temps, comme le phénomène fondamental de l'intellect humain, phénomène qui, par abstraction de son contenu émotionnel, donne le phénomène fondamental de la pensée mathématique, l'intuition de la dyade pure. Cette intuition de la dyade, intuition originaire des mathématiques, engendre non seulement les nombres un et deux, mais aussi tous les nombres ordinaux finis, attendu que l'un des éléments de la dyade peut être pensé comme une nouvelle dyade, et que ce processus s'itère indéfiniment; en poursuivant, il engendre le nombre ordinal infini le plus petit, ω. Finalement, cette intuition originaire des mathématiques, où s'unissent le connecté et le séparé, le continu et le discret, donne lieu immédiatement à l'intuition du continu linéaire, c'est-à-dire du "entre" qui ne se laisse pas épuiser par l'interposition de nouvelles unités, et qui donc ne peut jamais être pensé comme une simple collection d'unités».**

Brouwer met ainsi au jour ce qui existe de façon sous-jacente à une théorie de la représentation et qui en constitue sa base non-linguistique : l'intuition du continu, c'est-à-dire non pas seulement le continu pensé comme objet mais ce que nous avons appelé le rapport du continu à la continuité[93]. C'est ce rapport qui se donnait à voir dans le problème formaliste de la maîtrise du continu dans une représentation et qui apparaît ici comme donnée originaire. Cette intuition

---

[92] **L. E. J. Brouwer**, "Intuitionisme et Formalisme", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 43-44.

[93] Cf. Avant-propos, § 1.

originaire de la dyade pure, Brouwer la rapporte ensuite à la découverte du discontinu - d'une singularité - qui produit la distinction du sujet et de l'objet[94]. De ce point de vue, l'intuition du continu et du discret n'a pas à être rapportée, par la médiation de la représentation, à la faculté de pensée en général comme c'était le cas dans le formalisme : elle est la pensée elle-même. Dès lors, même si une psychologie reste possible, elle ne peut en rien rendre compte de l'intuition originaire et en reste à la pure extériorité de la pensée par rapport à elle-même, c'est-à-dire à ce qui en est linguistiquement descriptible[95]. Brouwer ne nie donc pas la possibilité d'une psychologie qui se limiterait à l'étude des êtres humains et des animaux en tant "qu'automates", c'est-à-dire à l'étude de ce qu'il est possible de modéliser du comportement en général par le biais d'algorithmes et qui serait comme un appendice à la physiologie. Mais, pour Brouwer, une telle psychologie laisse à l'extérieur de son domaine d'investigation ce qui constitue le moteur vital de la pensée, à savoir la capacité d'invention du sujet créateur, qui n'est pas susceptible, pour lui, d'être analysée en termes d'individus définis de façon solipsiste.

On vient de voir qu'un changement d'interprétation dans la notion de calcul introduisait un changement radical dans la façon de concevoir l'ontologie mathématique ainsi que le rôle de la pensée dans la constitution de l'objectivité mathématique. Du point de vue du projet d'intelligence artificielle, doit-on en conclure que la notion de calcul, dont la définition semble être universellement

---

[94] Brouwer l'appelle «la culpabilité volontaire de l'intellect». Cf. **L. E. J. Brouwer**, "Conscience, philosophie et mathématique", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 430 et 432.

[95] «Il ne peut y avoir qu'une psychologie de l'homme et de l'animal, qui, à titre d'extension de la physiologie, étudiera les organismes vivants automates, sans intellect ni libre arbitre. L'individu y entre aussi, nonobstant son rôle spécial de porteur du plaisir et de la souffrance, ainsi que des phénomènes qui accompagnent les émotions, les pensées, et les actions du sujet. Car en dépit de sa position dominatrice, le sujet occupe un champ de descriptibilité qui représente, en comparaison de celui de l'objet, une Cité du Vatican». **L. E. J. Brouwer**, "Conscience, philosophie et mathématique", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 429. Si l'on prend la comparaison finale de cette citation au sérieux, ce que le goût de Brouwer pour la tendance évocatrice propre à la langue naturelle semble imposer, il semble que ce qui constitue la sphère intime du sujet, inaccessible à la psychologie, relève, pour lui, de la catégorie du religieux.

acceptée, peut être soustraite à toute prise de position ontologique et utilisée de façon purement opératoire ? Nous ne le croyons pas : c'est *déjà* une prise de position ontologique que de considérer la notion de calcul comme purement opératoire. Il y a bien une *question du continu* dans le cadre de l'intelligence artificielle parce que la notion de calcul entretient avec celle de pensée formelle et celle de continu des liens étroits, comme l'ont montré les deux points de vue formaliste et intuitioniste. Il nous faut donc maintenant envisager de façon plus précise la notion de calcul et voir comment s'est constituée sa définition objective.

Chapitre II

Présentation classique
de la notion de calculabilité

1. Approche informelle du concept de calculabilité

Comme l'a montré notre analyse du débat entre formalistes et intuitionistes, la théorie du calcul exige que soit clarifiée la notion d'effectivité, qui se trouve être étroitement liée à celle d'algorithme. Jusqu'aux années 20-30, on n'avait pas thématisé la notion d'algorithme parce qu'on avait toujours su reconnaître l'existence d'un algorithme quand il s'en présentait un. Ce n'est qu'à partir du moment où l'on n'était plus assuré de trouver un algorithme pour un problème donné que s'imposa la nécessité de parvenir à définir la notion de façon générale.

11. Algorithme et fonction

Qu'entend-on, du point de vue général, par algorithme ? Il s'agit de l'énoncé d'une liste finie d'instructions devant être suivie selon un ordre donné. En suivant pas à pas la liste d'instructions, on doit aboutir au résultat après un nombre fini d'étapes, résultat que l'on sera capable de reproduire pour une infinité de cas particuliers qui seront tous traités de la même manière.

La notion d'algorithme était à l'origine cantonnée au domaine de l'arithmétique[96]. A partir de l'émergence de la théorie des fonctions au XVIII$^{\text{ème}}$

[96] On trouve des algorithmes de calcul dès l'antiquité grecque, comme par exemple le "crible d'Eratosthène" qui permet de trouver les nombres premiers.

siècle, la notion d'algorithme fut associée à la notion de fonction, qui avait elle-même le sens de "procédure de calcul" : à une valeur numérique de x correspondait, par une transformation effectuée par la fonction f, une valeur f(x). Mais le sens de la notion de fonction a progressivement évolué au cours du XIX^{ème} siècle jusqu'à signifier une correspondance quelconque entre éléments d'un ensemble de départ vers un ensemble d'arrivée sans que soit envisagée une procédure de calcul. Dès lors, il faut préciser le rapport entre fonction et algorithme en étudiant si, pour une fonction particulière d'entiers, il existe ou non un algorithme. Dans le cas où un algorithme existe pour la fonction en question, celle-ci est dite *calculable*.

La classe des fonctions calculables est donc une sous-classe de la classe des fonctions. Une des questions qui se pose naturellement à propos de cette sous-classe est de savoir comment la circonscrire, attendu qu'il est possible que l'on parvienne à exhiber un algorithme pour une fonction qui en était dépourvue jusqu'alors ou que l'on résolve le problème ouvert dont dépend le comportement de la fonction, comme c'est le cas aujourd'hui si l'on reprend l'exemple de Brouwer que l'on citait à l'instant, dans lequel le comportement d'une fonction dépend de la solution du théorème de Fermat. Les limites de cette sous-classe semblent mouvantes et de ce fait, partiellement indéterminées. Comment parvenir à les déterminer ?

12. Algorithme et décision

Quand on cherche à déterminer si un algorithme existe pour un problème ouvert donné, deux stratégies semblent praticables : on peut tenter soit, directement, de trouver l'algorithme en question, soit, indirectement, de démontrer qu'il ne peut pas exister. Mais il y a deux façons d'interpréter cette stratégie indirecte.

En effet, comme le montre le débat entre intuitionistes et formalistes, la découverte de l'absence d'un algorithme pour résoudre un problème ouvert donné peut être interprétée soit comme l'indice d'une limitation intrinsèque de notre pouvoir mathématique - ou tout au moins de l'existence d'une réalité qui ne cadre pas avec le pouvoir en question -, soit comme l'indice de l'existence d'un problème entièrement nouveau à résoudre, *celui de la délimitation de la classe des algorithmes*. En effet, la stratégie indirecte qui consiste à démontrer la non-existence d'un algorithme pour une question donnée peut être interprétée dans ce dernier cas comme exigeant que soit définie *exactement* la classe des algorithmes, cela en vue de montrer qu'aucun élément de cette classe n'est une solution pour la question posée[97]. Bref, dans cette interprétation, la question de la résolubilité par algorithme des problèmes ouverts exige de faire le détour par la définition de la classe des fonctions calculables. Seule cette interprétation de l'approche indirecte permet de poser la question des limites intrinsèques de la classe des fonctions calculables, puisqu'il faut déterminer la frontière entre ce qui se situe dans et hors de la classe en question. Étudions ces deux interprétations.

121. Effectivité et décision dans un contexte intuitioniste

Si l'on reprend l'exemple de la formule de calcul du nombre π, on se rend compte qu'il est possible de répondre au sujet de π à des questions du type : quelle est la 124ème décimale de son développement décimal ? ou : la 1245ème décimale du développement décimal de π est-elle le chiffre 2 ? Il suffit de poursuivre le développement jusqu'à la 124ème place pour répondre à la première question et de poursuivre le développement jusqu'à la 1245ème place puis de vérifier si le chiffre occupant cette place est bien 2 pour répondre à la seconde. Envisagée ainsi, la formule de calcul n'engendre pas seulement la suite des décimales correspondant au développement décimal du nombre mais devient une procédure logique de décision, c'est-à-dire un moyen de répondre par oui ou par non aux questions que l'on peut poser concernant le développement décimal du nombre examiné[98]. Bref, la notion mathématique d'algorithme permet d'aborder les questions logiques de décision.

---

97 Cf. **J. Mosconi**, *La constitution de la théorie des automates,* thèse de doctorat d'Etat, Université de Paris I, 1989, imprimé par l'Atelier National de reproduction des thèses, Université de Lille III, tome 1, p. 20-21 et **J. Largeault**, *La logique*, Presses Universitaires de France, Paris, 1993, p. 55.

98 Il s'agit d'une question portant sur l'appartenance, parce qu'on ne dit pas pourquoi la 124ème décimale du développement décimal de π est tel ou tel chiffre mais seulement lequel il est.

Est-on capable de répondre à toute question, par exemple sur le développement décimal d'un nombre réel comme π ? Non, car il y a des questions portant sur une propriété dont on ne connaît pas d'instance qui la vérifierait, sans avoir de preuve qu'il n'en existe aucune. Par exemple[99], on n'a aucun moyen de répondre à la question : "Le nombre de paires de chiffres consécutifs identiques dans le développement décimal de π est-il fini ou non ?". il faudrait, pour répondre à la question, connaître l'intégralité du développement décimal de π.

De façon générale[100], on se trouve donc dans la situation suivante : pour chaque objet *n* d'une classe infinie (comme par exemple les places du développement décimal de π), on sait déterminer si chaque objet *n* vérifie une propriété donnée P ou non (par exemple la propriété que deux chiffres consécutifs du développement décimal de π soient égaux ou pas), mais on n'a ni le moyen d'exhiber un objet vérifiant P (dans l'exemple choisi, de calculer que deux places consécutives du développement décimal de π sont occupées par le même chiffre) ni le moyen de démontrer que l'hypothèse que l'un des objets *n* de la classe infinie vérifie P implique contradiction (dans l'exemple, de démontrer qu'il ne peut pas y avoir deux chiffres égaux consécutifs dans le développement décimal de π).

Cette façon de traiter la question de la décision est de nature intuitioniste. Dans le contexte du débat entre intuitionistes et formalistes, cette façon d'interpréter la nature de la notion d'algorithme et la question de la décision revient en effet à formuler une critique du principe logique du tiers exclu. L'existence de problèmes non résolus, que leur absence de solution soit provisoire ou définitive, empêche de tenir à l'avance pour vraies des alternatives dont ni le membre affirmatif ni le membre négatif n'ont été prouvés vrais : le principe du tiers exclu n'a donc pas, dans ce cas, de validité universelle. L'existence de problèmes ouverts de ce type tend à corroborer l'idée selon laquelle il n'y a aucune raison *a priori* de considérer que tout problème mathématique est

---

[99] L'exemple est de Brouwer dans "Qu'on ne peut pas se fier aux principes logiques", traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], p. 22.
[100] Cf. **J. Largeault**, *Intuition et intuitionisme*, op. cit., p. 45.

soluble[101]. Dès lors, circonscrire une fois pour toutes la classe des algorithmes paraît impossible puisqu'on n'a pas les moyens qui permettraient de circonscrire la classe en question.

122. Effectivité et décision dans un contexte formaliste

Hilbert connaissait les exemples d'alternatives non décidables proposés par les intuitionistes : définir une partie d'ensemble infini par une propriété positive laisse en général indéterminé le complémentaire de cette partie[102]. Il a lui-même rappelé (suivant sans doute en cela les remarques critiques formulées par Brouwer) que de la fausseté de (n) P(n), on n'a pas le droit d'inférer qu'il existe un entier n tel que ¬ P(n). Mais, contrairement à Brouwer, il a cherché à obtenir des résultats généraux touchant la délimitation du domaine du décidable.

Pour ce faire, Hilbert souscrit à un principe philosophique très différent de celui de Brouwer, celui de la résolubilité universelle des problèmes mathématiques. La question de la décision ne peut se poser dans toute sa généralité qu'une fois adopté le principe en question. En effet, établir un résultat de décidabilité, qu'il soit positif ou négatif (c'est-à-dire un résultat d'impossibilité), exige dans les deux cas de supposer que le problème est quoi qu'il arrive résoluble (positivement ou négativement). Cependant, l'adoption d'un tel principe est nécessaire sans être suffisante. C'est par le bais de la logique que la question de la décidabilité prend toute sa portée : en effet, puisque les propositions mathématiques peuvent être représentées formellement sous l'aspect d'énoncés de la logique du premier ordre, la question de la décision touchant la dérivabilité au sein de cette logique vaut comme question de décidabilité pour tout énoncé mathématique[103]. Cette question est celle de l'*Entscheidungsproblem* au sein de la logique du premier ordre. On peut l'exprimer sous la forme suivante :

[101] Brouwer l'exprime sous cette forme : «Si chaque application du *principii tertii exclusi* en mathématiques accompagnait une procédure mathématique réelle, cela signifierait que chaque assertion mathématique (chaque assignation d'une propriété à une entité mathématique) pourrait être jugée, c'est-à-dire être soit prouvée soit réduite à l'absurde». **L. E. J. Brouwer**, *Cambridge Lectures on Intuitionism*, 1981, p. 5 cité dans **J. Largeault**, *L'intuitionisme*, op. cit., p. 40.

[102] Cf. **J. Largeault**, *Intuition et intuitionisme*, op. cit., p. 105.

[103] Cf. **J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit., p. 176.

peut-on décider pour tout énoncé formulé dans la logique du premier ordre s'il est ou non dérivable à partir des axiomes de cette logique ?

Pour parvenir à répondre à une telle question, il faut préciser ce que l'on entend par dérivabilité au sein de la logique. Or pour réussir à savoir s'il y a des limites à la dérivabilité, il faut réussir à préciser ce que l'on entend par calcul effectif. C'est en particulier une réponse négative à la question de l'*Entscheidungsproblem* qui obligerait à préciser quelles sont les limites à la dérivabilité puisque, dans ce cas, il y aurait comme un au-delà du dérivable.

C'est précisément cette analyse qu'opère Turing : en proposant une description du concept de calculabilité effective, celui-ci parvient à résoudre par la négative l'*Entscheidungsproblem*. Il faut donc maintenant en venir à la description de ce que l'on doit entendre par “calculable par machine”.

13. La notion de calculabilité

Turing répond à la question de la détermination de la classe des fonctions calculables en donnant un équivalent formel à l'expression de “calculable par algorithme” : toute fonction pour laquelle on a réussi à trouver un algorithme doit être calculable par “machine de Turing”. On aurait ainsi une correspondance entre d'une part les algorithmes et d'autre part les machines de Turing : si l'on possède un algorithme, on doit aussi posséder la machine de Turing qui lui correspond. Ainsi les fonctions calculables doivent-elles être “Turing-calculables”, c'est-à-dire calculables par “machine de Turing”. Bref, selon Turing, *quelle que soit l'étendue* de la classe des fonctions calculables, elles doivent toutes être “Turing-calculables” pour faire partie de la classe en question. On peut énoncer la thèse ainsi :

Thèse de Turing :

> **Toute fonction calculable par un être humain en suivant un algorithme peut être calculée par une machine de Turing.**

Il s'agit d'une thèse qui, en tant que telle, n'est pas susceptible de preuve : on ne peut pas en effet apporter de preuve à un énoncé qui met en rapport une notion à tout jamais informelle, celle de “fonction pour laquelle un être humain a trouvé un algorithme” et son équivalent formel, celui de “fonction pour laquelle il existe une machine de Turing”. La thèse ne permet donc pas d’exhiber un critère objectif d’appartenance à la classe des fonctions calculables. Il s’agit donc seulement d’une hypothèse permettant de vérifier l’appartenance d’une fonction particulière donnée à la classe des fonctions calculables mais qui n’est pas vérifiable globalement puisque l’on ne connaît pas toutes les fonctions calculables. Venons-en au concept de machine de Turing.

Quelle que soit la tâche remplie par une machine, on peut toujours interpréter sa table d’instructions comme représentant le *calcul d’une fonction* d’entiers à valeurs entières. Une fonction $\Phi(a)$ est dite Turing-calculable quand ses valeurs peuvent être calculées par une machine de Turing. On peut ainsi affirmer, grâce au formalisme de la machine de Turing, que la découverte d’un algorithme pour la résolution d’une classe donnée de problèmes est équivalente à la découverte d’une machine de Turing spécifique capable de fournir, dans un temps fini, la ou les solutions à la classe de problèmes en question. La difficulté qui demeure consiste seulement à établir, dans chaque cas, la correspondance entre la table d’instructions de la machine de Turing et l’algorithme.

Confronté à cette caractérisation de la notion de calculabilité, on peut avoir le sentiment que l’on a seulement “reculé pour mieux sauter” : en effet, il semble qu’*en plus* de la recherche d’un algorithme qui rendrait la fonction ou le nombre réel que l’on étudie calculable, il faut maintenant rechercher la “machine de Turing” qui correspond à cet algorithme. Mais il faut noter cependant, pour parer cette critique, que le rapport instauré entre les deux termes de la thèse de Turing *est en lui-même un rapport de représentation*, puisqu’il met en rapport un pôle subjectif et informel d’une part et un pôle objectif et formel d’autre part. Aussi cette façon de procéder revêt-elle, du point de vue d’une théorie de la représentation, un double avantage, puisqu’elle permet de faire porter l’attention et sur la caractérisation des fonctions calculables et sur la psychologie du mathématicien au travail, c’est-à-dire sur les termes objectif et subjectif de la représentation.

D'une part en effet, du point de vue objectif de la caractérisation des fonctions calculables, la thèse de Turing permet de considérer comme formant un tout des procédures de calcul qui sinon n'auraient en commun qu'un nom - celui d'algorithme. De ce point de vue, elle rend bien compte de l'essence de tout calcul.

D'autre part, d'un point de vue subjectif, en caractérisant par un trait commun tout algorithme, elle rend possible la constitution d'une étude de la psychologie du mathématicien au travail, étude qui sinon serait définitivement écartée. En effet, si l'on admet qu'il n'est pas possible de caractériser par un trait commun tout ce que l'on entend par procédure de calcul, alors on admet aussi que restent impénétrables les raisons psychologiques qui ont présidé à l'invention d'une nouvelle procédure et à sa reconnaissance en tant que procédure par la communauté des mathématiciens. C'est précisément pourquoi Brouwer, d'un même mouvement, pouvait déclarer qu'il n'y avait pas moyen de caractériser de façon univoque la notion de procédure de calcul légitime et que la notion de sujet créateur resterait impénétrable à la psychologie[104]. Mais ce n'est pas le cas si l'on se place du point de vue d'une théorie de la représentation.

L'optique adoptée par Turing a précisément pour but d'apporter un éclaircissement à la notion de fonction calculable et à celle de pensée du mathématicien. Bien plus, cette optique permet, comme nous allons le voir, d'*éclairer une notion par l'autre*. Cette approche de la notion de calculabilité ne serait récusée ni par les formalistes ni par les intuitionistes, malgré leur différence de principe et de pratique. Le concept de machine de Turing est en effet adéquatement défini, que l'algorithme représenté par la machine finisse par aboutir à un calcul achevé ou non et soit même achevable ou non. Bref, la notion de machine de Turing est susceptible de représenter les fonctions calculables, qu'elles soient totales ou partielles. Le désaccord entre intuitionistes et formalistes

---

[104] Tous les intuitionistes n'ont pas accepté l'attitude de Brouwer. Kreisel en particulier a proposé une axiomatique du sujet créateur en même temps qu'une caractérisation des procédures de calcul admises en intuitionisme. Cf. **G. Kreisel**, "Church's Thesis : a Kind of Reducibility Axiom for Constructive Mathematics" dans [**Myhill, Kino, Vesley eds.**, *Intuitionism and Proof Theory*, North-Holland, 1970] p. 121-150 cité dans **J. Largeault éd.**, *Intuition et Intuitionisme*, Vrin, Paris, 1993, pp. 84 et 101.

porte plutôt sur ce qui se situe “en dehors” du domaine de l’effectivité du calcul : la caractérisation de la réalité mathématique du point de vue de l’objectivité et celle de la psychologie du point de vue de la subjectivité.

Il nous faut donc maintenant exposer ce que l’on entend précisément par “machine de Turing”.

2. Présentation classique du concept de calcul par machine de Turing

Turing a construit le formalisme de la calculabilité par machine de Turing pour rendre compte formellement de l’expression de “procédure mécanique” telle qu’elle apparaît dans la définition informelle de la notion d’algorithme plutôt que pour rendre compte du terme de “machine” proprement dit. Plus exactement, le terme de mécanique, avant l’usage qu’en fait Turing, est interprété de façon purement métaphorique et possède la même signification que l’adjectif “servile”[105]. Turing envisage le terme de façon littérale[106]. Ces réserves faites, on peut donner une description de cette machine d’un genre nouveau que l’on appelle désormais en théorie de la calculabilité “machine de Turing”[107].

21. Description de la notion de machine de Turing

La machine de Turing est une machine “de papier”[108] ou encore un “automate abstrait” qui n’a rien à voir avec une machine matérielle.

211. La machine de Turing comme “boîte noire”

---

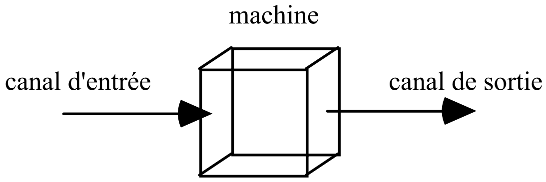
[105] Cf. **J. Mosconi**, *La constitution de la théorie des automates,* op. cit., tome 1, p. 19.

[106] Turing fait remarquer au § 2 de “Systems of Logic based on Ordinals”, (1939), Proceedings of the London Mathematical Society, ser. 2, vol. 45 : 161-228 reprint dans [**M. Davis ed.**, *The Undecidable* Raven Press, Hewlett, New York, pp. 154-222] : «Une fonction est dite “effectivement calculable” si ses valeurs peuvent être trouvées par quelque procédé purement mécanique (…). Nous pouvons prendre cet énoncé littéralement, en entendant par procédé purement mécanique un procédé qui pourrait être exécuté par une machine».

[107] L’expression n’est évidemment pas de Turing lui-même. On la trouve pour la première fois dans le compte-rendu de Church à l’article de Turing de 1936, publié dans le Journal of Symbolic Logic (2) (1937), pp. 42-43.

[108] L’expression est de Turing dans **A. M**. **Turing**, “Intelligent Machinery”, Executive Committee NPL, 1948, 1-20, Crown Copyright Reserved, H. M. S. O.; republié dans [*Key Papers : Cybernetics*, **C. R. Evans** et **A. D. G. Robertson eds.**, University Park Press, Manchester, England, 1968], p. 29.

Du point de vue d'un observateur extérieur, une machine de Turing apparaît comme une "boîte noire"[109] possédant un canal d'entrée et un canal de sortie. L'expression de boîte noire est justifiée par le fait que l'on ne précise pas comment la machine est physiquement mue ni comment s'agencent matériellement ses différentes parties. On ne prend en compte, sous cette description, que la nature de la transformation opérée entre le canal d'entrée et le canal de sortie sur les symboles fournis à la machine.



C'est un certain rapport particulier de transformation symboles d'entrée / symboles de sortie qui permet de caractériser en propre cette machine qu'est la machine de Turing. Une machine de Turing est en effet une machine qui transforme des symboles d'entrée en symboles de sortie en traversant une succession d'états discrets qui sont tous définissables à l'avance. Aussi la machine consiste-t-elle essentiellement en la mise en rapport de deux ensembles : d'une part un ensemble de symboles d'entrée et d'autre part un ensemble d'états de sortie qui définissent les actions de la machine.

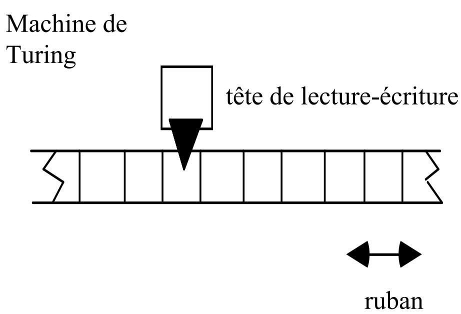
Plus précisément, une machine de Turing est une machine possédant une capacité de stockage externe qui se présente sous la forme d'un ruban de longueur infinie, divisé en cases sur lesquelles sont portés des symboles. La machine est dotée d'une tête de lecture-écriture capable d'observer le contenu des cases du

---

[109] L'expression est de Minsky dans **M. L. Minsky**, *Computation : Finite and Infinite machines*, Prentice-Hall, Inc. Englewood Cliffs, N. J. , 1967, p. 13; celui-ci fait remarquer : «L'utilisateur n'a pas normalement besoin de savoir ce qui se produit à l'intérieur de la boite. Aussi, sauf s'il est particulièrement intéressé par la compréhension de la façon dont la machine "travaille" ou dans sa modification, il doit seulement savoir quelles sont les propriétés "d'entrée-sortie". Quand on traite d'une machine de cette manière, nous l'appelons une "boite noire", indiquant par là que l'on se désintéresse du contenu».

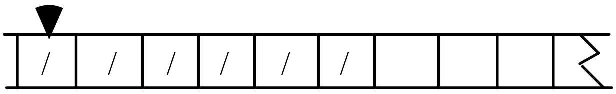
ruban, de se déplacer le long du ruban dans un sens ou dans un autre et de s'arrêter sur une case. Toutes les actions sont régies par une table d'instructions qui indique quelle action entreprendre - écriture ou mouvement. L'observation d'une case (sa lecture) peut se décomposer soit en effacement soit en écriture. A chaque moment discret du temps, moment que l'on peut indexer sur la suite des entiers naturels, la tête de lecture-écriture observe une case et une seule. Le couple formé de l'état interne de la machine à un moment t et de la case observée définit une configuration de la machine. La table d'instructions prescrit ainsi un comportement pour chaque configuration dans laquelle la machine peut se trouver. La machine effectue alors ce qui est prescrit par la table et produit un résultat. Ce mécanisme suffit à décrire la transformation qui affecte les symboles d'entrée pour en faire des symboles de sortie.

On peut alors représenter la machine de Turing sous cette forme :



212. Un exemple de calcul minimal

On peut prendre un exemple de calcul "minimal" pour décrire le fonctionnement d'une machine de Turing; par exemple, la machine qui consiste à rayer sur le ruban les symboles de barre transversale et à les remplacer par des symboles d'astérisque. Si l'on représente le ruban avec plusieurs symboles de barre transversale, des espaces avant et après la série de symboles et la tête de lecture-écriture sur la première case ayant un symbole de barre transversale pour contenu :



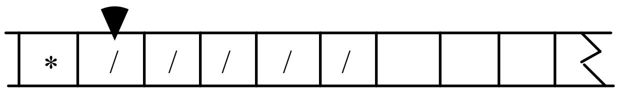
alors on peut remplacer les symboles de barre transversale par des symboles d'astérisque en utilisant une “table d'instructions” pour faire exécuter ce remplacement par une machine. On appelle “table d'instructions” d'une machine de Turing la table qui associe un acte à un couple (case lue, état interne)[110]. On peut alors présenter la table d'instructions sous la forme d'un tableau dans lequel la première colonne dénombre ses états et la première ligne décrit son alphabet :

	espace	/	*
1		*D	
2			

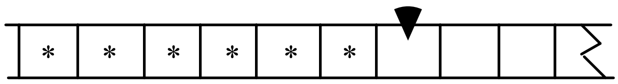
Chaque cellule du tableau peut recevoir trois types de symbole correspondant à trois paramètres, chacun optionnel. Le premier type correspond à ce qui va être écrit sur le ruban, le second, est soit “G” soit “D”, selon que la tête de lecture se déplace à droite ou à gauche; le dernier type est le nombre qui correspond au prochain état qui doit être exécuté. Dans l'exemple choisi, si la tête de lecture-écriture se trouve devant la première case ayant comme contenu une barre transversale, c'est-à-dire si l'état actuel est l'état n°1, alors quand le symbole de barre transversale est lu, la cellule active du tableau est celle qui contient “*D”. Donc une astérisque est écrite à la place de la barre transversale et la tête de lecture se déplace à droite. Puisqu'il n'y a pas de troisième paramètre,

[110] Comme le fait remarquer Jean Mosconi : «Le tableau d'une machine de Turing quelconque schématise dans une métalangue à l'usage de l'observateur la façon dont est réglé un certain comportement, ou plus précisméent un certain traitement de symboles». **J. Mosconi**, *La constitution de la théorie des automates,* op. cit., tome 1, p. 44.

l’état suivant est encore l’état n°1. Le ruban apparaît alors comme ceci :



Quand la machine a écrit une astérisque sur la dernière barre transversale, la tête de lecture se trouve sur une case vide. Comme on n’a pas spécifié d’instruction quand la tête de lecture se trouve dans l’état n°1 sur une case vide, la machine s’arrête faute d’instructions. Le ruban apparaît alors ainsi :

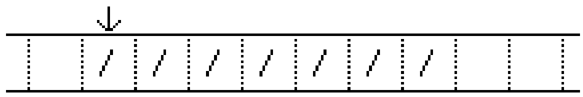


Le calcul est achevé.

213. Un exemple de calcul sans arrêt

On peut construire une machine de Turing pour un algorithme dont la fonction serait d’intervertir les symboles de barre transversale et d’astérisque. Cette machine, comme on va le voir, ne s’arrête pas mais intervertit indéfiniment les deux types de symboles. Malgré cette absence d’arrêt, la machine de Turing qui effectue ce calcul possède une table d’instructions parfaitement définie. Comme nous l’avions déjà remarqué, la notion d’arrêt d’une machine de Turing pour telle entrée n’est donc pas nécessaire à la définition précise de la notion de calcul par machine de Turing.

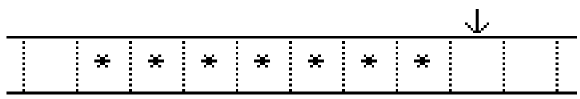
Dans l’exemple choisi, le ruban et la tête de lecture se présentent sous la forme :



La table d’instructions comporte deux états et deux symboles d’alphabet :

	espace	/	*
1	G2	*D	
2	D1		/G

Quand la tête de lecture-écriture rencontre “/”, elle est dans l’état n°1. Selon les instructions contenues dans la cellule colonne 2 ligne 2, elle écrit “*” et se déplace d’une case à droite. Elle réitère la même opération jusqu’à ce qu’elle soit arrivée au bout des cases remplies à droite. Elle change le dernier “/” en “*” et se déplace d’une case à droite : elle tombe sur une case vide.



Elle retourne alors à gauche d’une case et entre dans l’état n°2. Sur la première case remplie par une “*”, elle écrit “/” et se déplace d’une case à gauche, selon les instructions de la cellule colonne 3 ligne 2. Elle réitère cette opération jusqu’à parvenir à l’extrême gauche, c’est-à-dire jusqu’à ce qu’elle tombe sur une case vide: elle retourne à droite et entre dans l’état n°1, etc.

Le calcul ne s’arrête pas et la table d’instructions est parcourue indéfiniment.

214. Un exemple de calcul numérique avec arrêt : la fonction successeur

Pour les deux exemples que l’on vient de décrire, c’est presque de façon métaphorique que l’on parle de calcul. C’est en effet par extension que l’on en parle quand il n'est pas fait usage de nombres dans l'alphabet utilisé par la machine que l'on cherche à décrire. Pour parvenir à utiliser des nombres et à effectuer ainsi un calcul proprement dit, il faut réussir à représenter les nombres entiers dans l'alphabet.

Prenons l'exemple de la fonction successeur[111], que l'on appellera la fonction Φ. Le problème consiste à trouver une machine, appelée T, qui calcule cette fonction, c'est-à-dire qui soit capable, quand on lui donne un nombre n quelconque en entrée, d'écrire le nombre n + 1 en sortie.

Supposons que l'on représente le nombre 1 par deux traits "||" que l'on écrit chacun sur une case du ruban et que l'on insère (à un moment 0) le bout du ruban dans la machine T en plaçant la tête de lecture sur la case portant un symbole et se situant le plus à droite. Si la table d'instructions de la machine T a été bien rédigée, la machine doit, après un nombre fini d'états correspondant à un nombre fini de moments discrets, parvenir à inscrire trois traits consécutifs sur trois cases adjacentes séparées par des cases vides, avant de s'arrêter.

Si l'on se donne la liberté de représenter la case vide par 0 et la case pleine par 1 et si l'on ajoute aux deux symboles "d" et "g" un nouveau symbole "c" pour indiquer que la tête de lecture reste sur la même case, on peut dresser la table d'instructions de la machine T nécessaire pour effectuer le calcul de la fonction successeur de 1 :

[111] J'emprunte l'exemple à Kleene dans **S. C. Kleene**, "Turing's Analysis of Computability and Major Applications of It", dans [**Herken R. ed.**, *The Universal Turing Machine*,, Oxford Science Publications, Oxford University Press, 1988],p. 24-25.

	0	1
1	0c0	1d2
2	0d3	1d9
3	1g4	1d3
4	0g5	1g4
5	0g5	1g6
6	0d2	1d7
7	0d8	0d7
8	0d8	1d3
9	1d9	1g10
10	0c0	0d11
11	1c0	1d11

Au moment 0, la tête de lecture est placée sur la case la plus à droite contenant un symbole et la machine a été placée dans l'état 1 : elle lit alors le contenu de la case, trouve un trait, se déplace à droite et entre dans l'état 2, comme l'indique le tableau au premier rang, deuxième colonne.

On voit que le tableau ne représente pas la liste des moments nécessaires à l'effectuation du calcul de la fonction successeur mais seulement la liste des états de la machine dans lesquels elle peut entrer, au besoin plusieurs fois, pour réussir à effectuer le calcul. Pour le calcul du successeur de 1, il faudrait décrire 23 moments successifs pour parvenir à ce que la machine s'arrête après avoir écrit dans un suite de trois cases (puisque 1 est représenté par deux traits) le successeur de 1. Si on représente les moments à gauche, ce qui est écrit sur le ruban à droite et les états par un exposant, on a le tableau suivant :

<u>Moment</u>	<u>Ruban</u>
0	$0 \; 1 \; 1^{1}$
1	$0 \; 1 \; 1 \; 0^{2}$
2	$0 \; 1 \; 1 \; 0 \; 0^{3}$
3	$0 \; 1 \; 1 \; 0^{4} \; 1$
4	$0 \; 1 \; 1^{5} \; 0 \; 1$
5	$0 \; 1^{6} \; 1 \; 0 \; 1$
6	$0 \; 1 \; 1^{7} \; 0 \; 1$
7	$0 \; 1 \; 0 \; 0^{7} \; 1$
8	$0 \; 1 \; 0 \; 0 \; 1^{8}$
9	$0 \; 1 \; 0 \; 0 \; 1 \; 0^{3}$
10	$0 \; 1 \; 0 \; 0 \; 1^{4} \; 1$
11	$0 \; 1 \; 0 \; 0^{4} \; 1 \; 1$
12	$0 \; 1 \; 0^{5} \; 0 \; 1 \; 1$
13	$0 \; 1^{5} \; 0 \; 0 \; 1 \; 1$
14	$0^{6} \; 1 \; 0 \; 0 \; 1 \; 1$
15	$0 \; 1^{2} \; 0 \; 0 \; 1 \; 1$
16	$0 \; 1 \; 0^{9} \; 0 \; 1 \; 1$
17	$0 \; 1 \; 1 \; 0^{9} \; 1 \; 1$
18	$0 \; 1 \; 1 \; 1 \; 1^{9} \; 1$
19	$0 \; 1 \; 1 \; 1^{10} \; 1 \; 1$
20	$0 \; 1 \; 1 \; 0 \; 1^{11} \; 1$
21	$0 \; 1 \; 1 \; 0 \; 1 \; 1^{11}$
22	$0 \; 1 \; 1 \; 0 \; 1 \; 1 \; 0^{11}$
23	$0 \; 1 \; 1 \; 0 \; 1 \; 1 \; 1^{0}$

Le dernier moment parvient à écrire l'état inactif après avoir écrit trois 1 consécutifs séparés par une case vide : la machine est parvenue à calculer le successeur du nombre 1.

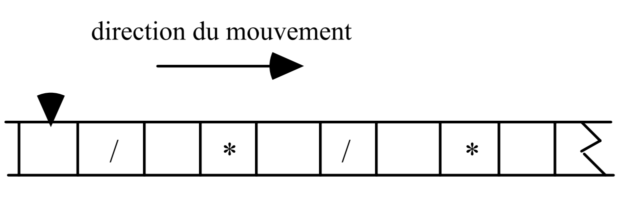
215. Un exemple de calcul numérique sans arrêt : le calcul d'une suite

On peut construire une machine qui calcule la suite infinie 0101010101 … .

En partant d'un ruban vide, la table d'instructions de la machine qui calcule la suite en question est des plus simples puisqu'il suffit de construire une table d'instructions composée de quatre états définis seulement pour une case vide. Si, par commodité, on veut garder le même vocabulaire que dans les exemples qui précèdent l'exemple de la fonction successeur, et que l'on représente 0 par "/" et 1 par "*", on a le tableau suivant :

	espace
1	0D2
2	D3
3	1D4
4	D1

La machine imprime, en séparant chaque numéral par une case vide, la suite / * / * / * / * … qui correspond à la suite 010101 … :



Le calcul de la suite ne s'arrête pas puisque le ruban est vide et qu'il n'y a donc pas de couple (case lue, état interne) dans lequel la case lue serait remplie.

Ces exemples permettent de décrire la notion de machine de Turing d'un point de vue général.

216. Description de la notion générale de machine de Turing

Tout calcul effectué par machine de Turing est décomposable dans les éléments que l'on vient de mentionner : l'alphabet de symboles utilisés en entrée et en sortie, le déplacement de la tête de lecture-écriture, l'état de la machine à un moment *t* du temps et éventuellement, l'arrêt de la tête de lecture-écriture sur une case du ruban. Détaillons ces divers éléments.

Un alphabet est composé d'une liste finie S de symboles (s_1, …, s_m).

Du point de vue de la fonction d'entrée, chaque case du ruban de la machine peut recevoir un symbole d'entrée choisi dans la liste S, ou rester vide (le symbole est alors s_0). La liste finie des symboles de l'alphabet est donc (s_0, …, s_m).

La fonction de sortie de la machine est double : elle consiste à écrire sur la case observée et à se déplacer le long du ruban. Aussi la réponse de la machine est-elle composée de deux éléments : le premier est un symbole qui est écrit sur la case observée (il peut s'agir du même symbole que le symbole qui vient d'être lu) et qui est tiré du même ensemble alphabétique que les symboles d'entrée (s_0, …, s_m); le deuxième est l'un des deux symboles "G" (signifiant : "déplacer la tête à gauche") et "D" (signifiant : "déplacer la tête à droite"). On doit donc rajouter ces deux symboles d'instruction à l'alphabet de la machine.

Cette fonction de sortie définit un état de la machine. A chaque moment discret du temps, la machine se trouve dans un état soit actif, soit passif. Les états actifs sont ceux dans lesquels la machine est en mouvement; ils font partie d'une liste Q finie q_1, … q_k; l'état passif est l'état q_0, dans lequel la machine s'arrête. Il y a donc une liste finie d'états de sortie q_0, … q_k.

L'alphabet, composé d'une part des symboles d'instructions nécessaires à la transition entre états et d'autre part des symboles d'entrée et de sortie, est écrit sur le ruban. Celui-ci est infini et on peut le concevoir comme infini dans deux directions ou bien comme borné à gauche ou à droite. Dans toutes les expositions cependant, seul un nombre fini de cases est rempli : aussi quand la machine démarre, le ruban doit-il être vide excepté pour un nombre fini de cases. A chaque étape, le ruban est donc en fait fini, mais on peut rajouter autant de cases que l'on veut, selon les besoins du calcul effectué par la machine, d'où l'expression de "ruban infini".

A chaque cycle d'opérations, la machine démarre dans un des états q, lit l'un des symboles de S écrits sur la case, écrit sur la même case le nouveau symbole tiré des symboles de S, se déplace à gauche ou à droite et entre alors dans un nouvel état tiré des états de Q.

On peut ainsi rendre compte[112] du fonctionnement d'une machine de Turing en faisant remarquer que son comportement peut être décrit par un ensemble de quintuplets de la forme : (1) ancien état q_i; (2) symbole lu s_j; (3) nouvel état q_{ij}; (4) symbole écrit s_{ij}; (5) direction du mouvement d_{ij}. Les relations entre les éléments de ces quintuplets peuvent être décrites au moyen de trois fonctions : la fonction N qui détermine le nouvel état; la fonction F qui détermine le nouveau symbole à écrire; la fonction C qui détermine de quel côté la tête de lecture-écriture doit se déplacer sur le ruban. On a alors :

nouvel état $q_{ij} = N\ (q_i, s_j)$

symbole écrit $s_{ij} = F\ (q_i, s_j)$

direction du mouvement $d_{ij} = C\ (q_i, s_j)$

Cet ensemble de quintuplets suffit à définir les relations entre les états de la machine, ses entrées et ses sorties. La liste de ces quintuplets constitue la table d'instructions de la machine et permet ainsi de donner une description complète de la structure logique de la procédure exécutée par la machine.

---

112 J'emprunte la description à Minsky dans **M. L. Minsky**, *Computation : Finite and Infinite machines*,op. cit., p. 119-120.

La description de la notion générale de machine de Turing peut être complétée par un appendice : certaines machines de Turing ont une propriété particulière qui revêt un certain intérêt pour qui tente de déterminer le champ du calculable. Il s'agit des machines de Turing dites "universelles".

22. Description de la notion de machine de Turing universelle

Les machines de Turing dites "universelles" sont des machines de Turing comme les autres, au sens où elles sont composées des mêmes éléments. A quoi tient leur "universalité" ? Au fait qu'elles peuvent *imiter* le fonctionnement de n'importe quelle autre machine de Turing. L'universalité de ces machines de Turing provient donc de leur universelle capacité d'"imitation". Que faut-il entendre par "imitation" dans le contexte des machines de Turing ? On dit qu'une machine de Turing A imite une machine de Turing B quand A, munie des instructions adéquates, est capable *d'effectuer* le calcul effectué par B. L'imitation est universelle quand on reconnaît à la machine A la capacité d'imiter le comportement de n'importe quelle machine.

Quel peut être l'intérêt de cette classe particulière de machines que sont les "machines universelles" ?

221. Intérêt de la notion de machine de Turing universelle pour une théorie de la représentation

Pratiquement, on ne voit pas exactement quel pourrait être l'intérêt de ce type de machines. En effet, du point de vue de celui qui construit une machine en rédigeant sa table d'instructions, la table d'instructions d'une machine de Turing universelle ne peut être que plus compliquée que la machine qu'elle imite puisqu'il faut rajouter à sa table d'instructions les instructions spécifiques qui lui permettent d'imiter la machine à imiter. Or ceci apparaît bien comme une difficulté supplémentaire. De même que lorsqu'on a caractérisé la notion d'algorithme par le biais de la notion de machine, de même ici on a l'impression de "reculer pour mieux sauter" et de compliquer la recherche d'un algorithme pour un problème donné au lieu de le simplifier : maintenant, non seulement la recherche d'un algorithme passe par la recherche de la machine de Turing éventuelle qui lui correspondrait mais il faut rajouter à cette recherche les instructions permettant à une machine universelle d'imiter la machine que l'on aura réussi à construire en vue d'effectuer l'algorithme. Ce n'est donc pas du strict point de vue de la caractérisation de la notion de calcul que la notion de machine universelle possède un intérêt.

Si l'on se rappelle en revanche que la notion de machine de Turing a été élaborée dans le contexte de la "thèse de Turing" selon laquelle toute fonction calculable par un être humain en suivant un algorithme est calculable par une machine de Turing, on se rend compte que la notion de machine universelle revêt une grande importance : la simulation qui définit en propre la notion de machine de Turing universelle est le point de vue par lequel tout calcul, quelle que soit sa complexité, peut être envisagé d'un point de vue fini. Essayons de comprendre pourquoi.

Selon le schéma de la thèse de Turing, on pourrait supposer que le calcul de nombres réels très compliqués exige d'augmenter le nombre d'états d'esprit du calculateur humain nécessaires à l'effectuation du calcul, nombre qui finit par tendre vers l'infini à mesure que la complexité s'accroît. Dans ce cas, le calcul peut-il être effectué par une machine de Turing ? En fait, il suffit de rajouter des symboles sur le ruban de la machine pour contourner cette difficulté : plus le calcul est compliqué, plus le nombre d'instructions à écrire sur le ruban peut être élevé, comme le fait remarquer Turing lui-même[113]. Le cas de la machine universelle présente précisément une confirmation de ce fait[114] : la machine

---

[113] Cf. **A. M. Turing**, (1936), 'On Computable Numbers with an Application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society*, 42 : 230-265, § 9; republié dans [**M. Davis ed.**, *The Undecidable* Raven Press, Hewlett, New-York, 1965, p. 115-154], p. 136 : «Si nous admettions une infinité d'états d'esprit, certains seraient "arbitrairement proches" et seraient confondus. Encore une fois, cette restriction n'affecte pas sérieusement le calcul, puisque l'utilisation d'états d'esprit plus compliqués peut toujours être évité en écrivant plus de symboles sur le ruban».

[114] C'est ce que remarque J. Webb dans son introduction au texte de Gödel, "Some remarks on the undecidability results", 1972a, reproduit dans **K. Gödel**, [*Collected Works*, t. II, Oxford University Press, Oxford, 1990], p. 300, dans laquelle il renvoyait à un texte de Turing : «Ainsi la complexité de la machine à imiter est concentrée sur le ruban et n'apparaît en aucune façon dans la machine universelle elle-même». **A. M. Turing**, (1947), 'Lecture to the London Mathematical Society', 20

universelle sert à contourner l'aspect infini du nombre d'états d'esprit nécessaires au calcul des nombres réels pris globalement puisqu'elle permet de simuler, grâce à une table de configurations finie, tout calcul effectué par une machine en ne prenant en compte que le numéro de code correspondant à chacune d'entre elle. L'aspect universel présent dans la notion de simulation possède dès lors un intérêt du point de vue de la confirmation *psychologique* qu'elle apporte à la thèse de Turing.

Le calculateur humain, selon l'algorithme qu'il veut mettre en pratique, construit en effet telle ou telle machine de Turing. D'un point de vue psychologique, le calculateur utilise toujours le même ressort pour effectuer cette correspondance : pour tel algorithme, utiliser telle table d'instructions. La correspondance entre un algorithme et la table d'instructions fait donc l'objet d'une procédure générale. Cette procédure générale de mise en correspondance ne pourrait-elle pas être elle-même opérée par une machine ? C'est possible, puisqu'on peut concevoir des machines de Turing dites "universelles" qui ont la particularité d'opérer la mise en effectuation, si leur table d'instructions est correctement rédigée, de n'importe quel algorithme. Il suffirait ainsi de définir une machine de Turing ayant cette caractéristique pour réussir à effectuer n'importe quel calcul, comme un calculateur humain unique est capable de s'adapter à chaque problème particulier et de trouver à la fois l'algorithme correspondant au problème qu'il se pose, la table d'instructions de la machine de Turing qui correspond à l'algorithme et leur mise en rapport.

Aussi, du point de vue d'une théorie de la représentation, la notion de machine universelle permet-elle de serrer au plus près le rapport entre le pôle subjectif et le pôle objectif de la représentation, tel que ce rapport s'exprime dans la thèse de Turing. En effet, de même qu'un même calculateur humain est capable de rechercher différents algorithmes pour résoudre les différents problèmes qu'il rencontre, de même la machine universelle est capable, selon les instructions qui lui sont confiées, de calculer ce que différentes machines de Turing peuvent

---

february 1947; republié dans [**B. E. Carpenter** et **R. W. Doran eds.**, *A. M. Turing's ACE report of 1946 and other papers*, MIT Press, Cambridge, Massachussetts, 1986, pp. 106-124], p. 112.

calculer.

La notion de machine universelle permet donc de préciser la signification de la thèse de Turing. Revenons un instant sur celle-ci; elle s'énonçait sous la forme : "Toute fonction calculable par un être humain en suivant un algorithme peut être calculée par une machine de Turing". La thèse proposait, grâce au concept de machine de Turing, un équivalent formel à la notion informelle de "calculable par un être humain" et se présentait comme une implication logique : si une fonction est calculable par un être humain alors elle est calculable par une machine de Turing. Une fois l'existence de machines de Turing universelles mise au jour, on peut préciser la thèse et l'exprimer sous la forme : "Toutes *les* fonctions calculables par un être humain en suivant un algorithme peuvent être calculées par *une* machine de Turing universelle". Dans cette formulation, on prend comme un fait acquis la caractérisation du champ du calculable et on met l'accent sur le fait qu'il est possible de donner un équivalent formel non pas seulement aux cas particuliers de recherche d'une machine de Turing mais à l'instance de recherche des algorithmes elle-même, à savoir l'être humain qui calcule. Cette analogie entre machine universelle et calculateur humain est cependant quelque peu trompeuse : alors que le calculateur humain *trouve*, par des moyens intuitifs, un algorithme, la machine universelle le *reçoit* en entrée sur son ruban. Nous reviendrons sur cette distinction.

On voit cependant d'ores et déjà l'intérêt que revêt cette dernière formulation puisqu'elle permet de caractériser plus avant le pôle subjectif du rapport de représentation - à savoir la notion psychologique de calculateur -, en en proposant un équivalent formel, celui de machine universelle. Bref, c'est surtout du point de vue de la théorie de la représentation que la notion de machine universelle possède un intérêt puisqu'elle permet de préciser ce qui sinon serait resté dans l'ombre, à savoir l'instance subjective qui calcule et non pas seulement les manifestations de cette instance.

Après cette description très générale du rôle joué par la notion de machine de Turing universelle du point de vue de la théorie de la représentation, il nous faut déterminer plus avant les caractères propres à ce type de machines.

222. Remarques sur les deux traits propres à l'imitation

On a vu que l'universalité de certaines machines de Turing provenait de leur capacité à imiter tout calcul effectué par machine de Turing. On peut donc décrire le rapport d'imitation comme un rapport entre une "machine imitante" et une "machine imitée". Comment s'opère cette imitation ?

Deux traits permettent de caractériser celle-ci : d'une part, la capacité à recevoir les instructions d'une autre machine sur son propre ruban et d'autre part la capacité à obéir effectivement aux instructions copiées, c'est-à-dire à effectuer le calcul proprement dit. On va voir que ces deux caractéristiques renvoient aux deux pôles du rapport de représentation tel qu'il apparaît dans la thèse de Turing : la première caractéristique, la réception, renvoie à l'instance psychologique mise en action dans la recherche d'un algorithme tandis que la seconde, l'effectuation, renvoie à ce que la machine exécute objectivement.

222. 1. La capacité à recevoir les instructions d'une autre machine

On a vu à l'instant qu'il fallait opérer une distinction entre d'une part l'instance psychologique du calculateur qui, par des moyens intuitifs, *trouve* un algorithme et la machine de Turing qui lui correspond et d'autre part une machine de Turing universelle qui *reçoit* sous la forme d'un numéro de code les instructions d'une autre machine. Cette distinction permet d'éclairer ce qu'il faut entendre par la "capacité de réception" d'une machine de Turing universelle. En quoi consiste cette "réception" ?

En fait, du point de vue du fonctionnement d'une machine de Turing universelle, il n'y a pas à proprement parler de capacité de "réception" qui serait différente de la capacité à recevoir des instructions en entrée sur son ruban : la "réception" des instructions d'une autre machine se présente sous la même forme que n'importe quelle instruction de n'importe quelle machine. Une machine de Turing universelle reçoit en effet les instructions d'une autre machine sous la forme "ordinaire" d'un codage indexé sur les entiers : à tel numéro de code correspond telle machine de Turing. Comme il est possible de coder les entiers naturels sur un ruban de machine de Turing, il est possible d'opérer ce codage des machines de Turing en une liste infinie dénombrable. On peut alors concevoir une relation fonctionnelle entre les tables d'instructions des deux machines, la description de la machine imitée jouant le rôle d'un argument de fonction pour la machine imitante. C'est cette relation fonctionnelle qui confère à la machine imitante la particularité d'imiter n'importe quelle machine. On peut alors démontrer qu'il existe au moins une machine de Turing qui peut imiter n'importe quelle machine de Turing, c'est-à-dire qui peut calculer n'importe quelle fonction calculable.

Une machine de Turing universelle n'est donc pas dotée d'une caractéristique supplémentaire tangible qui lui conférerait sa spécificité de machine universelle. C'est plutôt le *mathématicien* qui *reconnaît* dans un certain montage d'une table d'instructions cette capacité à l'imitation universelle. De quelle nature est cette reconnaissance ?

Le mathématicien reconnaît à une machine de Turing la capacité à recevoir *virtuellement* les tables d'instructions de n'importe quelle machine sans qu'il lui soit évidemment possible de vérifier si cette "réception" est bien possible pour chaque numéro de code. On se trouve dans la même situation que lorsque l'on reconnaît à la fonction successeur la capacité virtuelle de pouvoir calculer le successeur de n'importe quel entier sans effectuer ce calcul : la suite des entiers étant infinie, on accorde à la règle de succession un statut universel sans vérifier la vérité de cette règle au cas par cas, puisque cette vérification envelopperait l'infini. La capacité de réception de la machine de Turing universelle renvoie donc en fait à la capacité du mathématicien de coder les tables d'instructions de toutes les machines de Turing qu'il peut parvenir à concevoir. On retrouve donc ici ce qui faisait la spécificité de la thèse de Turing : l'appel à une instance psychologique comme fondement de la possibilité d'un calcul. La capacité de "réception" propre à la description de la notion de machine de Turing universelle permet de caractériser le champ d'investigation propre à l'intuition : il s'agit d'un champ de nature *virtuelle* que l'intuition explore au moyen d'algorithmes.

222. 2. La capacité à effectuer le calcul d'une autre machine

A la capacité purement virtuelle consistant à pouvoir recevoir les instructions d'une autre machine de Turing sur son propre ruban, vient s'ajouter la capacité, pour une machine de Turing universelle, à effectuer les opérations successives telles qu'elles sont décrites dans la table d'instructions de la machine à imiter. Il s'agit là d'une capacité *effective* et non pas seulement virtuelle qui renvoie au pôle objectif de la thèse de Turing : c'est en effet quand une machine de Turing en imite une autre en effectuant les opérations de cette machine que l'on peut s'assurer que la première machine est bien candidate à être une machine universelle. Aussi est-ce l'effectuation qui confère le caractère d'universalité à telle ou telle machine de Turing. C'est pourquoi la "réception" d'une table d'instructions sur le ruban de la machine universelle apparaît comme une condition nécessaire mais non suffisante à la construction d'une machine de Turing universelle. Quels rapports entretiennent alors les deux caractéristiques ? Pourrait-on imaginer séparer entièrement les deux traits caractéristiques de la notion de machine de Turing universelle et inventer un autre type de machine qui posséderait la capacité de recevoir la table d'instructions d'une machine sans en effectuer le calcul ?

Il semble possible de concevoir un autre type de machine de Turing qui aurait la capacité de recevoir sur son ruban les instructions codées de tables d'instructions de machine de Turing mais qui, au lieu de les exécuter, leur ferait subir une transformation quelconque. Mais en fait, on se rend compte que la "réception" n'a d'intérêt que si elle est suivie par une effectuation, que celle-ci consiste à exécuter les instructions ou à les transformer d'une manière ou d'une autre. Bref, la virtualité d'une "réception" ne prend son sens que par rapport à une effectuation : le champ du virtuel doit se rabattre sur celui de l'effectif pour que, de façon rétroactive, on puisse considérer la possibilité d'autres "réceptions" possibles. "Réception" et effectuation semblent donc aller de pair et la première n'a d'intérêt que parce qu'elle implique la seconde.

23. Le virtuel et l'effectif dans la notion de machine de Turing

Les remarques que l’on vient de faire touchant les rapports des deux traits caractéristiques de l’imitation dans le cas des machines universelles possède une portée générale dans la mesure où elles permettent de caractériser le comportement de toute machine de Turing. On peut en effet présenter par ce biais un problème de portée générale touchant le comportement de toute machine de Turing, que l’on appelle le “problème de l’arrêt” : peut-on savoir à l’avance si tout calcul aura ou non une fin ? Autrement dit : peut-on réussir, d’un point de vue complètement général, à caractériser *dans le virtuel* le résultat d’un calcul sans avoir à l’exécuter ? Il s’agit bien d’un problème général parce qu’il permet de tracer des limites à la calculabilité.

Cette question revient en effet à s’interroger sur l’existence d’une machine de Turing (ou d’un algorithme) qui serait capable de résoudre le problème de l’arrêt pour toute machine de Turing sur une entrée (ou “input”) donnée. S’il existait une telle machine, il existerait une machine de Turing capable de connaître *globalement* le comportement de chaque machine de Turing (c’est-à-dire le résultat de l’effectuation du calcul de la machine : son arrêt ou son absence d’arrêt) à partir de leur aspect *local* (c’est-à-dire à partir de la simple inspection du contenu de leur table d’instructions). On peut démontrer que la réponse au problème de l’arrêt est négative : une telle machine de Turing, capable de séparer radicalement le champ virtuel de la réception des instructions et le champ effectif de l’effectuation du calcul, n’existe pas.

231. La solution négative au problème de l’arrêt

La démonstration d’une solution négative au problème de l’arrêt ne fait pas intervenir la seconde caractéristique de la machine universelle : on part de l’hypothèse que nous formulions à l’instant, à savoir qu’il peut y avoir une machine qui sépare le virtuel de l’effectif, contrairement à la machine universelle, et on montre qu’elle conduit à une contradiction[115].

---

[115] Cette façon de présenter la démonstration est dûe à Minsky. Cf. **M. L. Minsky**, *Computation : Finite and Infinite machines*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1967, p. 147-148.

Problème de l'arrêt :

Y-a-t-il une procédure de décision qui dirait pour chaque machine de Turing T ayant t pour ruban et pour tout input (liste de symboles) sur ce ruban, c'est-à-dire pour une machine de Turing quelconque, si la machine s'arrête ou pas ?
On raisonne par l'absurde en partant de l'hypothèse qu'une telle procédure existe et on tente de décrire la machine de Turing qui lui correspondrait, comme le veut la démarche propre à la "thèse de Turing"

Hypothèse :
Supposons qu'une procédure de décision existe pour toute machine de Turing
Alors la machine de Turing qui lui correspondrait devrait avoir la forme suivante :

On part d'une machine de Turing quelconque T

Machine T

Nature de T :
machine de Turing quelconque

But de T :
la machine T lit les instructions inscrites sur son ruban t et les exécute

Résultat atteint par T :
T parvient ou ne parvient pas à un arrêt selon l'input qui lui est soumis

T écrit en
output :

OUI

Si T s'arrête
pour t donné

NON

Si T ne s'arrête
pas pour t donné

Le comportement de la machine T

peut être imité par une machine capable

d’exécuter la procédure de décision

Quelle forme aurait cette machine ?

Machine D

Nature de D :

machine imitante et “décisionnelle”

But de D :

en tant que machine imitante, elle lit son ruban (t, d_T),

où t est le ruban de la machine T et dt la description de la machine T

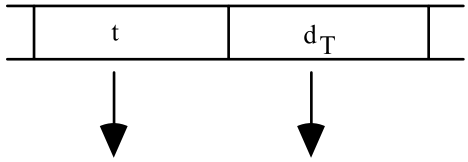
Résultat atteint par D :

en tant que machine “décisionnelle”, D s’arrête ou pas selon que T s’arrête ou pas,

quelle que soit la machine T ayant t pour ruban.

Dans les deux cas, l’output de D vaut comme décision

Description de D ayant t comme input :



Cet ensemble de cases correspond à la partie du ruban de D contenant l'input t

Cet ensemble de cases correspond à la partie du ruban de D contenant la description (de n° de code "d_T ") de la machine T ayant t pour ruban

D écrit en output :

OUI

et D s'arrête Si T s'arrête pour t

NON

et D ne s'arrête pas Si T ne s'arrête pas pour t

<u>Cas particulier</u> :

La machine D fonctionne aussi pour des cas de type (d_T, d_T) où la partie “input” est égale à la partie “description de machine”, c’est-à-dire pour un ruban contenant les instructions nécessaires à une description du fonctionnement de T plutôt que pour un ruban dont l’input permettrait d’exécuter une tâche quelconque.

Les instructions (d_T, d_T) en input à la machine D rendent possible l’interprétation du fonctionnement de T comme fonctionnant sur un entier qui est l’explicitation de son propre fonctionnement

Description de D ayant comme d_T input :

d_T	d_T

Du point de vue de
l'observateur extérieur

Cet ensemble de cases correspond à la partie du ruban de D
contenant la description du ruban (de n° de code "d_T ")

Cet ensemble de cases correspond à la partie du ruban de D contenant la description (de n° de code "dt") des instructions nécessaires au fonctionnement de T

Du point de vue
de la machine D :

Les numéros de code décrivant un
ruban et la machine sont les mêmes

D écrit en output :

OUI

et D s'arrête
Si T s'arrête
pour l'input
d_T

NON

et D ne s'arrête pas
Si T ne s'arrête pas
pour l'input d_T

On peut concevoir une autre machine E

qui serait capable d'assurer la

décision sur un seul input d_T et non sur (d_T, d_T)

Machine E

<u>Nature de E</u> :

machine dupliquante et "décisionnelle"

<u>But de E</u> :

En tant que machine dupliquante, quand elle lit l'input d_T,

elle le duplique en (d_T, d_T)

Le couple (d_T, d_T) peut alors servir au fonctionnement de D

<u>Résultat atteint par E</u> :

En tant que machine "décisionnelle", la machine E parvient à une décision

concernant la possibilité de l'arrêt sur un seul input d_T :

E s'arrête sur d_T si T s'arrête sur d_T

et E ne s'arrête pas sur d_T, si T ne s'arrête pas sur d_T

Description de E ayant d_T comme input :

$\boxed{d_T} \longrightarrow \boxed{\text{dupliquer}} \longrightarrow \boxed{d_T \mid d_T} \longrightarrow D$

E écrit

en output :

OUI

et la machine (de numéro de code d_T) s'arrête sur l'input d_T
(Si D s'arrête
Alors E s'arrête)

NON

et la machine (de numéro de code d_T) ne s'arrête pas sur l'input d_T
(Si D ne s'arrête pas
Alors E ne s'arrête pas)

On peut concevoir une autre machine E* semblable à E

mais dont l’une des sorties comporte

une boucle qui empêche tout arrêt

Machine E*

Nature de E*:

La machine E* n’est ni imitante, ni dupliquante, ni “décisionnelle”

But de E* :

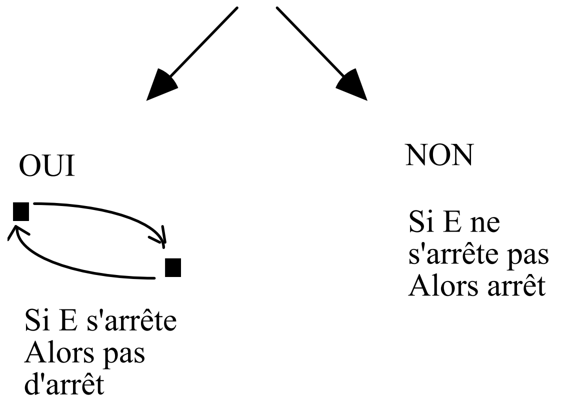
Elle lit d_T

Résultat atteint par E* :

E* s’arrête si E ne s’arrête pas sur l’input d_T

et E* ne s’arrête pas si E s’arrête sur l’input d_T

Description de E* ayant d_T comme input :



Que se passe-t-il dans le cas
où l'on donne à E* en input sa propre description ?
d_E* joue le rôle que jouait d_T dans le cas précédent

<u>Résultat atteint par E*</u>
<u>avec sa propre description</u> d_E* <u>comme input</u> :

Si on donne comme input d_E* à la machine E*
Alors E* s'arrête sur d_E* ssi E* ne s'arrête pas sur d_E*:

Contradiction

Il n'y a donc pas de machine E*, ni de machine E, ni de machine D
Il n'y a donc pas de machine "décisionnelle" qui pourrait déterminer,
pour toute machine de Turing sur n'importe quel input,
si la machine va s'arrêter ou pas sur cet input.

La démonstration a donc consisté, en jouant sur la différence entre virtuel et effectif, à exhiber *une* machine pour laquelle une prise de décision serait contradictoire. De ce point de vue, la démonstration a montré que le problème de l'arrêt doit être résolu dans le virtuel et donc qu'il n'est pas soluble. Quel sens revêt ce résultat quand on cherche à préciser la relation entre les aspects virtuels et effectifs dans la notion de machine de Turing ?

232. Le virtuel et l'effectif dans la thèse de Turing

Maintenant que l'on a montré en quel sens on pouvait concevoir une limite à la calculabilité sur le cas particulier du problème de l'arrêt, on peut préciser les rapports qu'entretiennent le virtuel et l'effectif au sein de la thèse de Turing. L'étude de ces rapports vont permettre de préciser ce que l'on entend par calcul et par intuition.

232. 1. L'effectivité et la notion de calcul

A quoi reconnaît-on, dans une manipulation symbolique, la présence d'un calcul effectif ?

Notre analyse a montré que, d'un point de vue intuitif, la notion d'effectivité n'était pas liée à l'aspect achevé du calcul : ce n'est pas parce qu'un calcul n'a pas de fin qu'on ne le reconnaît pas en tant que calcul, et ce, ni dans le cas de la notion informelle d'algorithme, ni dans le cas de sa traduction formelle de machine de Turing. En effet, si le concept de calcul ou de procédure effective est psychologiquement intuitionnable même sans arrêt, alors ce n'est pas son caractère entièrement déterminé jusqu'à un arrêt, qui fait, psychologiquement, la force du concept de calculable. Aussi peut-on souscrire à la remarque de Wang Hao qui reproduit l'opinion de Gödel [116]:

> **«Gödel fait remarquer que la notion précise de procédures mécaniques est clairement mise en lumière par les machines de Turing produisant des fonctions récursives partielles plutôt que générales. Autrement dit, la notion intuitive ne requiert pas qu'une procédure mécanique doive toujours s'arrêter ou réussir. Une procédure quelque fois mise en échec, si elle est définie clairement, est encore une procédure, c'est-à-dire une façon bien déterminée de mener à bien une effectuation».**

Si ce n'est pas l'aspect achevé qui permet de préciser ce que l'on entend par effectivité du calcul, alors comment caractériser celle-ci ? Comme l'a montré la description des caractéristiques de la notion de machine de Turing, c'est par

---

[116] **W. Hao**, *From mathematics to philosophy*, Routledge and Kegan Paul, London, 1974, p. 84.

rapport à un plan d'intelligibilité purement virtuel que l'on parvient à caractériser l'effectivité du calcul. Plus précisément, c'est l'articulation des deux plans où sont manipulés des symboles, l'un virtuel - renvoyant à l'intuition du mathématicien - et l'autre effectif - renvoyant à une manipulation finie - qui permet de caractériser la notion d'effectivité. C'est cette articulation du virtuel et de l'actuel qui constitue le fond de la thèse de Turing et qui permet d'assurer le passage entre une notion informelle et sa traduction formelle. On peut donc dire que la notion de calcul, telle qu'elle est constituée par l'intuition, permet en retour de caractériser l'intuition comme cette faculté qui se manifeste localement dans l'effectif et globalement dans le virtuel. Cette articulation peut être précisée.

232. 2. L'intuition, entre virtuel et effectif

Notre analyse a permis de dégager cinq traits permettant de décrire les rapports qu'entretiennent les notions d'intuition, d'algorithme et de machine dans le cadre de la thèse de Turing.

Description de l'activité de recherche

Premièrement, la thèse de Turing a posé que la faculté d'intuition est à tout jamais *informelle*, parce que la notion d'algorithme l'est aussi.

Deuxièmement, la description de la machine de Turing a posé que, pour l'intuition humaine, il y a identité entre la recherche d'un algorithme et la recherche d'une machine.

Il n'y a donc pas de “supériorité” de l'intuition sur la machine : du point de vue de la thèse de Turing, l'idée d'une “supériorité” ou d'une “infériorité” de l'intuition par rapport à la machine n'a d'ailleurs aucun sens puisque la notion de machine est un *mode d'expression* de l'intuition au même titre que la notion d'algorithme. En revanche, cela ne veut pas dire qu'il faille identifier *sous tous les points de vue* l'intuition et ses modes d'expression, algorithme ou machine.

Description de l'acte de calcul

Troisièmement, l'acte de calcul est identifié dans l'intuition humaine et dans la machine.

L'acte de calcul est le point de vue sous lequel on peut subsumer l'intuition et la notion de machine. Il faut donc distinguer entre l'*activité* de recherche d'un algorithme ou d'une machine et l'*acte* de calcul proprement dit. De ce point de vue, le fait que l'intuition humaine puisse rechercher des algorithmes et des machines qui leur correspondent alors que la machine ne fait que recevoir des instructions ne doit pas entrer en ligne de compte pour ce qui est de la description de l'*acte* effectif de calcul.

Caractérisation négative de l'activité de recherche

Quatrièmement, la solution négative au problème de l'arrêt a montré qu'il y avait une impossibilité logique à trouver un algorithme ou une machine pour résoudre le problème en question.

La thèse de Turing apparaît ici de façon négative : il n'y a pas d'algorithme pour résoudre le problème de l'arrêt, donc il n'y a pas de machine, donc la recherche intuitive d'un algorithme ou d'une machine est vouée à l'échec quoi qu'il arrive. L'intuition apparaît ici comme aussi démunie que la machine.

Caractérisation positive de l'activité de recherche

Cinquièmement, la machine de Turing universelle permet de préciser la différence entre acte de calcul et activité de recherche.

Pour ce qui est de l'acte de calcul, le cas de la machine de Turing universelle a montré que tout calcul peut être effectué par une machine unique. Aussi, de même que le calculateur humain devient à l'occasion un multiplicateur quand le problème qu'il se pose exige d'opérer une multiplication, de même une machine universelle devient multiplicatrice si elle reçoit sur son ruban le numéro de code d'une machine qui multiplie[117]. De ce point de vue, il y a identité entre les démarches de l'intuition et de la machine.

Pour ce qui est de l'activité de recherche, selon ce qui a été exposé dans les points 1 et 2, l'intuition humaine est capable de chercher *dans le virtuel* un algorithme ou une machine alors que la machine universelle en *reçoit* une description toute faite sur son ruban. Il n'y a donc pas identité entre intuition et machine universelle sur ce point, *sans qu'il y ait différence radicale entre elles* puisque l'intuition peut se trouver dans un cas où la recherche d'un algorithme ou d'une machine n'aboutirait à rien. C'est donc seulement dans le moment qui va de la recherche d'un algorithme (ou d'une machine) à sa découverte que l'intuition se distingue de la machine, puisque sitôt cette découverte faite, une machine pourrait effectuer le calcul rendu accessible par l'algorithme (ou la machine). Ce moment, entièrement virtuel, ne peut jamais apparaître comme tel puisqu'il ne peut se manifester qu'au sein de l'actualité effective du calcul, c'est-à-dire au sein du domaine où il y a identité entre intuition et machine. C'est ce que remarquait Gödel quand il décrivait les rapports de l'intuition et de la machine en s'appuyant sur qu'il considérait être un résultat absolument démontré concernant leur rapport [118] :

> **«1. L'esprit humain est incapable de formuler (ou de mécaniser) toutes ses intuitions mathématiques. C'est-à-dire : S'il a réussi à formuler l'une d'entre elles, ce fait lui-même produit une nouvelle connaissance intuitive, par exemple la consistance de ce formalisme. Cet état de fait peut être appelé l'incomplétibilité des mathématiques. D'autre part, en se fondant sur ce qui a été prouvé jusqu'à maintenant, il reste possible qu'il puisse exister (et même qu'il soit empiriquement possible de découvrir) une machine à prouver des théorèmes qui en réalité *soit* équivalente à l'intuition humaine mais ne puisse *prouver* qu'elle le soit, ni même qu'elle puisse prouver qu'elle produise seulement des théorèmes exacts de la théorie finitaire des nombres».**

Il reste à montrer, du point de vue d'une théorie de la représentation, que la thèse de Turing possède bien un contenu objectif, c'est-à-dire qu'elle décrit au mieux le concept de calculabilité. Pour ce faire, il faut réussir à montrer la

---

117 J'emprunte l'exemple à J. Mosconi dans **J. Mosconi**, *La constitution de la théorie des automates,* op. cit., tome 1, p. 44.

118 Ces réflexions de Gödel furent exposées à la 25ème Josiah Willard Gibbs Lecture qui se tient à Providence le 26 décembre 1950. Elles sont reproduites dans le chapitre "X. 7. Gödel on minds and machines" de **W. Hao**, *From mathematics to philosophy*, op. cit., p. 324.

correspondance entre la notion de machine de Turing et d'autres notions mathématiques permettant également de définir le concept de calculabilité.

3. Récursivité et machine de Turing

Le concept de machine de Turing n'est en effet pas le seul moyen de donner un aspect formel à la notion d'algorithme et de caractériser ainsi la notion de calculabilité. Historiquement, d'autres mathématiciens ont mis au jour à peu près en même temps[119] d'autres concepts qui visent le même but. On s'est rendu compte alors qu'il était possible de prouver que ces formulations étaient toutes équivalentes entre elles. De ce point de vue, la multiplicité des formulations et la preuve de leur équivalence a eu pour effet "psychologique" de corroborer, dans l'esprit des mathématiciens, l'idée selon laquelle les définitions formelles qu'ils proposaient de la notion d'algorithme étaient exhaustives. Pour réussir à mesurer la portée générale de ce résultat, un détour par la notion de fonction récursive, mise au jour par Gödel sur une suggestion de Herbrand[120], semble approprié.

On peut réussir à décrire les nombres et les fonctions calculables à partir d'une analyse menée en termes entièrement arithmétiques sans passer par le concept de machine de Turing. En effet, tous les mouvements d'une machine de Turing, c'est-à-dire toutes les transformations de ses états internes, peuvent être exprimés par le biais d'un certain nombre de fonctions arithmétiques et de l'opération de composition de fonction qui sont définies indépendamment du concept de machine de Turing.

31. Fonctions récursives primitives

Voici comment on définit la classe des fonctions récursives primitives : on se donne un ensemble de fonctions initiales et un ensemble d'opérations qui

---

[119] Entre 1934 et 1936, trois formulations visant à définir la notion de calculabilité ont été proposées : celle de Church-Kleene (λ-calcul) en 1932-1934, celle de Herbrand-Gödel (fonctions récursives) en 1934 et celle de Turing (1936-1937). D'autres formulations sont possibles, comme celle, plus récente, de Conway (jeu de la vie).

[120] Historiquement, le formalisme des fonctions récursives a précédé celui de Turing de deux ans, puisque Gödel l'a présenté, sur une suggestion faite auparavant par Herbrand, en 1934.

permettent d'obtenir à partir d'un ou de plusieurs éléments déjà formés de la classe des fonctions récursives primitives un nouvel élément[121].

Les fonctions initiales sont les suivantes : 1°. La fonction successeur notée S qui à un nombre x fait correspondre son successeur $x + 1$ dans la suite des entiers; 2°. L'ensemble C des fonctions constantes type C_a^n cette notation désignant la fonction qui fait correspondre à un n-uple $<x_1, \ldots, x_n>$ l'élément quelconque a donné de l'ensemble des entiers; 3°. L'ensemble U des fonctions "projection" U_i^n où pour chaque i et chaque n tels que $1 \leq i \leq n$, la fonction est définie par $U_i^n (x_1, \ldots, x_i, \ldots, x_n) = x_i$ pour toutes les valeurs de $x_1, \ldots, x_i, \ldots, x_n$ (par exemple la fonction U_2^4 telle que $(x_1, x_2, x_3, x_4) = x_2$).

Les opérateurs sont les suivants : 1°. L'ensemble infini Ω_C des opérateurs de composition; Chaque opérateur $U_C^{m,n}$ permet d'obtenir, pour m et n donnés supérieurs à 0, à partir de m fonctions $\chi_1, \ldots, \chi_m$ de n variables chacune et d'une fonction ψ de m variables, la fonction ϕ de n variables telle que $\phi (x_1, \ldots, x_n) = \psi (\chi_1((x_1, \ldots, x_n), \ldots, \chi_m ((x_1, \ldots, x_n))$ 2°. L'ensemble infini Ω_R des opérateurs de récurrence. Pour l'opérateur U_R^n, dans le cas où $n = 0$, on a $\phi (0) = a$; $\phi (S(y)) = \chi (y, \phi (y))$.

On peut montrer à partir de ce formalisme qu'un grand nombre de fonctions usuelles sont récursives primitives. Par exemple, la fonction +, définie par les deux équations : 1°. $+(x, 0) = x$; 2°. $+(x, S(y)) = S (+ (x,y))$ est récursive primitive.

La classe des fonctions récursives primitives n'est cependant pas suffisante pour regrouper toutes les fonctions calculables. On utilise pour le montrer un argument de "diagonalisation", qui produit une fonction calculable sans pourtant être récursive primitive[122]. L'argument a cette forme : en

---

[121] J'utilise ici la présentation de R. Martin dans **R. Martin**, *Logique contemporaine et formalisation*, Presses Universitaires de France, Paris, 1964, pp. 107-108.

[122] Historiquement, cette argumentation semble due à Ackermann, dans **W. Ackermann** "On Hilbert's Construction of the real numbers", traduction anglaise dans [**J. van Heijenoort ed.**,

établissant par exemple la liste effective des fonctions récursives primitives d'une variable, notée ϕ_0, ϕ_1, ϕ_2, …, ϕ_n, … et en prenant en considération la fonction ψ définie de la façon suivante : quel que soit n, $\psi(n) = \phi_n(n) + 1$. Cette fonction est certainement calculable puisque le numéro d'ordre de ϕ assure qu'en partant de n on peut parvenir à calculer la fonction ϕ_n. Mais ψ n'est pas récursive primitive parce que, si elle l'était, elle occuperait un rang déterminé m dans la liste des fonctions récursives primitives et serait identique à la fonction ϕ_m. On aurait donc : quel que soit n, $\phi_m(n) = \phi_n(n) + 1$, ce qui pour $n = m$ entraîne $\phi_m(m) = \phi_m(m) + 1$. On conclut par l'absurde que ψ est à la fois calculable et non récursive primitive.

32. Fonctions récursives générales

Il faut alors envisager un élargissement de la classe des fonctions récursives primitives. Cet élargissement permet de définir les fonctions récursives générales, qui sont constituées à partir de la classe des fonctions récursives primitives et l'ajout d'un nouvel ensemble infini d'opérateurs, noté Ω_μ, qui désigne l'opérateur dit de minimalisation. On définit l'opérateur de la façon suivante : Si ψ est une fonction de n + 1 variables telle que $(x_1),\ldots,(x_n)\ (m\,y)\ [\psi(x_1, \ldots, x_n, y) = 0]$, alors l'opérateur Ω_μ^{n+1} permet d'obtenir à partir de ψ la fonction ϕ de n variables, telles que $\phi(x_1, \ldots, x_n) = \mu y\ (\psi(x_1, \ldots, x_n, y) = 0)$ où μy signifie "le plus petit y tel que (…)", aucune borne supérieure n'étant fixée à y.

Contrairement au cas des fonctions récursives primitives, les fonctions récursives générales ne garantissent pas l'existence d'un résultat pour toute valeur que peut prendre leur argument, parce qu'il n'y a pas moyen de savoir de manière effective si, dans le cas de l'opérateur Ω_μ, il existe ou non un y qui remplisse la condition de la définition. Il est possible qu'il faille égrener la liste infinie des

---

From Frege to Gödel : a source book in mathematical logic, Harvard University Press, Cambridge, 1967], p. 493-507.

entiers pour tenter de trouver une valeur à *y*.

Il faut alors distinguer, au sein des fonctions récursives générales, les fonctions récursives partielles définies pour certaines valeurs de l'argument et les fonctions récursives totales définies pour toutes leurs valeurs. Mais cette distinction a pour conséquence de retrouver, sous une nouvelle formulation, le résultat négatif du problème de l'arrêt telle que nous l'avions déjà rencontré dans le cadre du calcul par machine de Turing.

En effet, on peut énumérer et constituer une liste des fonctions récursives partielles puisqu'il est possible de savoir si toutes les opérations utilisées dans l'élaboration d'une récursion sont effectives sans avoir à se demander si le calcul s'arrête ou pas. On peut alors énumérer récursivement les fonctions récursives partielles (de même que l'on pouvait supposer l'existence d'une machine de Turing universelle capable de simuler à elle toute seule tout calcul effectuable par une autre machine de Turing). Mais il n'y a pas moyen de savoir quelles fonctions récursives partielles sont en fait des fonctions récursives totales. Dès lors, les fonctions récursives totales ne peuvent pas être énumérées et il n'est pas possible de savoir *a priori* si une fonction récursive partielle est ou non définie pour la valeur d'un de ses arguments (de même qu'il n'y avait pas de solution "décisionnelle" pour le problème de l'arrêt pour toutes les procédures effectuables par machines de Turing).

33. Exhaustivité de la caractérisation de la notion de calculabilité

Les différentes traductions de la notion de calculabilité corroborent, sans jamais la prouver, l'exhaustivité des définitions de la notion de calculabilité. C'est surtout sur son exhaustivité que l'on insiste habituellement, en laissant de côté la caractérisation de la faculté psychologique d'intuition qu'il est possible de faire à partir d'elle. C'est pourquoi on présente habituellement la notion de calculabilité sous l'aspect d'une pure et simple *définition*. C'est ce qui se produit dans les présentations classiques de la notion de calculabilité.

Emil Post réagit contre cette tendance en faisant remarquer qu'exprimer une thèse sous la forme d'une définition revient à occulter son aspect informel, aspect qui manifeste en fait l'existence d'une limitation interne aux pouvoirs mathématiques des êtres humains, dans la mesure où cette thèse exige d'être continuellement reconfirmée[123] :

> **«Mais masquer l'identification derrière une définition occulte le fait qu'une découverte fondamentale concernant les limitations du pouvoir mathématique de *Homo Sapiens* a été réalisée et nous rend aveugle au besoin de sa continuelle vérification»**

Cet avertissement énoncé par Post doit nous mettre en garde contre la tendance à forger des définitions à l'apparence objective. Si l'on en revient à l'énoncé de la thèse de Turing, on voit que celle-ci vise tout d'abord à déterminer clairement - et non pas formellement - ce que l'on entend par calculabilité. Pour ce faire, Turing se place dans une optique philosophique bien précise, celle d'une théorie de la représentation. C'est dans le cadre de cette théorie philosophique qu'il parvient à produire un certain nombre de résultats mathématiques. Il nous faut donc aborder maintenant la façon dont Turing a exposé originellement ses résultats.

---

[123] Cf. **E. Post** (1936), "Finite Combinatory Processes - Formulation I", Journal of Symbolic Logic, I, réédité dans [**M. Davis ed.**, *The Undecidable*, Raven Press, Hewlett, New York, 1965], pp. 288-291.

Chapitre III

La notion de calculabilité chez Turing : mathématique, logique et psychologie

La présentation classique de la notion de calculabilité et tout particulièrement la façon dont on expose habituellement la démonstration du problème de l'arrêt, laissent de côté un certain nombre de points qui ont leur importance quand on cherche à préciser la portée du rapprochement entre la théorie de la calculabilité et le modèle mécanique du fonctionnement de l'esprit, rapprochement qui a donné naissance au projet d'intelligence artificielle.

Trois points méritent d'être particulièrement soulignés concernant la présentation originelle de la théorie de la calculabilité par machine telle qu'elle est exposée par Turing. Celui-ci a tout d'abord[124] développé la notion de machine dans un contexte mathématique, pour rendre compte de la question de la calculabilité des nombres réels. Il a, dans un deuxième temps[125], envisagé la même question d'un point de vue logique quand il s'est penché sur l'application de son modèle mécanique du calcul au cas de la dérivabilité dans les systèmes axiomatiques. De ces points de vue différents découlent deux types de remarques psychologiques sur le rapport entre intuition humaine et fonctionnement mécanique. Il nous faut donc commencer par exposer les deux points de vue envisagés par Turing, le premier mathématique et le second logique, avant d'en

---

[124] Cf. **A. M. Turing**, (1936), 'On Computable Numbers with an Application to the Entscheidungsproblem', Proceedings of the London Mathematical Society, 42 : 230-265; republié dans [**M. Davis ed.**, *The Undecidable* Raven Press, Hewlett, New-York, 1965, p. 115-154].

[125] Cf. **A. M. Turing,** (1939), "Systems of Logic based on Ordinals", Proceedings of the London Mathematical Society, ser. 2, vol. 45 : 161-228 reprint in [**M. Davis ed.**, *The Undecidable* Raven Press, Hewlett, New-York], p. 154 sq.

venir à la description du modèle de l’esprit qui en découle.

1. Le point de vue mathématique adopté par Turing

La formulation classique des résultats touchant la notion de calculabilité par machine de Turing laisse de côté ce qui relie le point de vue de Turing au débat sur le fondement des mathématiques tel qu’on l’a exposé au chapitre 1. Ce débat avait eu pour principal résultat de mettre l’accent sur la controverse entre formalistes et intuitionistes concernant la nature du continu et les moyens de l’appréhender. Or c’est précisément sur les moyens à notre disposition pour exhiber les éléments du continu que porte l’analyse de Turing dans son article de 1936 puisqu’il se pose la question de savoir ce qui fait que l’on considère un nombre réel comme calculable. Aussi la perspective mathématique adoptée par Turing permet-elle de préciser ce qui est déterminable par le calcul au sein du continu et d’aborder ainsi, au moins indirectement, la question de sa nature ultime. Le cas plus spécifiquement logique de l’*Entscheidungsproblem* - qui n’a pas en lui-même de rapport direct avec la question de la détermination du continu - n’est invoqué qu’à titre d’“application”, comme l’indique d’ailleurs le titre complet de l’article de 1936, “On Computable Numbers with an Application to the *Entscheidungsproblem*”[126].

11. La calculabilité et les nombres réels

Comme on vient de le remarquer, Turing, en se posant la question de savoir comment préciser la notion de calculabilité, se place d’emblée dans le système des nombres le plus général, celui des nombres réels. Le problème qu’il aborde peut, en première analyse, s’exprimer sous la forme de la question suivante : comment réussir à caractériser par le calcul une collection infinie ? Plusieurs cas doivent être distingués.

---

[126] Comme le remarque très fermement A. Hodges, le concept de machine de Turing dépasse largement le cadre technique du problème de la décision tel qu’il avait été posé par Hilbert : «L’essence du résultat de Turing fut la découverte d’un concept ayant *une application en* logique et qui avait ses racines dans des idées se trouvant hors du champ des mathématiques». **A. Hodges**, “Alan Turing and The Turing Machine” in [*The Universal Turing machine*, **R. Herken ed.**, Oxford Science Publications, Oxford University Press, 1988], p. 5.

111. Position du problème

Dans le cas des entiers naturels, on possède à la fois la représentation d'exemplaires particuliers de nombres comme 1, 3 ou 7 et une opération, l'opération successeur, qui permet de former un entier naturel quelconque[127].

Dans le cas des nombres réels, on possède la représentation d'exemplaires particuliers de ces nombres comme $\sqrt{2}$ ou π mais on ne possède pas d'opération qui permettrait de caractériser un nombre réel quelconque[128]. La notion de nombre réel quelconque fait donc difficulté parce qu'il semble qu'il n'y a aucun moyen de caractériser de façon homogène tous les nombres réels. C'est cette difficulté qui rend nécessaire la distinction du continu géométrique et du continu arithmétique : dans le cas de la droite géométrique en effet, tous les points qui la composent sont génériquement homogènes alors que ce n'est pas le cas des nombres réels, puisque seuls certains d'entre eux sont accessibles par le calcul. Une question qu'il paraît naturel de se poser est donc celle des moyens grâce auxquels on peut circonscrire, parmi la classe des nombres réels, ceux qui sont accessibles par le calcul. C'est cette question qui définit la problématique générale de la notion de calculabilité telle qu'elle est envisagée par Turing.

112. Analyse du problème

Une fois que l'on a réussi à se former la représentation d'un exemplaire de nombre réel par le biais d'une intuition géométrique (c'est le cas par exemple de $\sqrt{2}$ ou π), comment peut-on l'approcher par le biais du calcul ? La caractérisation

---

[127] Cf. **J. Largeault**, *Intuition et Intuitionisme*, Vrin, Paris, 1993, pp. 90-91.

[128] La définition d'une opération permettant de caractériser les réels est la difficulté majeure de l'arithmétique transfinie dont l'opération fondamentale serait celle de mise à la puissance. Cf. **K. Gödel**, "What is Cantor's continuum problem ?", Bull. Am. Math. Soc., vol. 58, 1952 et dans [**P. Benacerraf** et **H. Putnam**, *Philosophy of mathematics,* CUP, Cambridge, 2nd edition, 1983], traduction française dans [**J. Largeault éd.**, *Intuitionisme et théorie de la démonstration*, op. cit.], pp. 516-517 : «Ainsi le problème du continu se révèle être une question issue de la "table de multiplication" des nombres cardinaux, à savoir le problème d'évaluer un certain produit infini (en fait le problème non-trivial le plus simple qui puisse se trouver dans cet ordre d'idées)».

d'un nombre réel peut être introduite de plusieurs façons[129], mais nous nous limiterons à celle qui aborde l'introduction des nombres réels par le biais de leur développement décimal, parce que c'est celui que privilégia Turing pour rendre compte de la notion de calcul et que c'est son optique qui fait l'objet de notre analyse.

Tout nombre réel est caractérisé par son développement décimal mais il n'est pas toujours possible de définir ce développement par le biais d'une équation. Seuls les nombres réels qui peuvent être définis par des équations algébriques ou transcendantes peuvent être exprimés par leur développement décimal. En général, le développement décimal d'un nombre réel n'est ni fini ni périodique, contrairement à celui des nombres rationnels; mais dans le cas où l'on peut déterminer une suite définie de manière effective qui converge vers le nombre réel en question, on peut à bon droit considérer que ce nombre réel est *calculable*. Le développement décimal peut alors être donné par une formule permettant de calculer le nombre, c'est-à-dire d'un algorithme de calcul. Par exemple, π peut être défini par "$\pi = 4\ (1- 1/3 + 1/5 - \dots)$". En se donnant un temps infini, il devient en droit possible de calculer les unes après les autres les décimales du développement de π. Le caractère infini du développement décimal fait qu'en pratique, il est exclu de calculer des places de décimales trop grandes, mais qu'elles restent en droit calculables.

La notion de machine de Turing représente, comme on l'a vu au chapitre précédent, l'analogue formel de la notion d'algorithme telle qu'elle est évoquée ici à propos du cas de π. Cependant, la notion d'algorithme et sa contrepartie formelle peuvent être envisagées sans que soit évoqué le cas particulier du calcul d'un nombre réel. Bien plus, le cas du calcul de l'image d'un nombre réel par une fonction calculable est en fait plus complexe que le cas le plus immédiat, à savoir celui de la définition d'une fonction calculable sur des entiers. Turing lui-même

---

[129] Il y en a au moins cinq : (1) par des emboîtements d'intervalles d'extrémités rationnelles; (2) comme classes d'équivalences de suites de Cauchy; (3) par des coupures dans les rationnels; (4) par des développements infinis de fractions décimales ; (5) de façon intuitionniste, comme déploiement finitaire conçu comme *species* de suites convergentes de nombres rationnels. Cf. **J. Largeault**, *Intuition et Intuitionisme*, op. cit., pp. 149-151.

remarque au paragraphe 10 de “On Computable Numbers …” :

> **«Nous ne pouvons définir les fonctions calculables d’une variable réelle, puisqu’il n’y a pas de méthode générale permettant la description d’un nombre réel, mais nous pouvons définir une fonction calculable d’une variable calculable».**

Pourquoi Turing a-t-il choisi le cas du calcul des nombres réels alors qu’il aurait pu envisager le cas plus simple du calcul des fonctions d’entiers ?

113. Démarche suivie par Turing

En fait, le cas du calcul d’un nombre réel est particulièrement expédient pour le but que s’est fixé Turing et qui est de parvenir à définir la notion de calculabilité. On sait en effet depuis Cantor, grâce en particulier à son argument de diagonalisation, que les nombres réels sont en nombre infini non-dénombrable[130]. Or, pour parvenir à tracer des limites à la calculabilité, il faut réussir à trouver un cas où celle-ci est prise en défaut. De ce point de vue, le cas des nombres réels s’impose naturellement puisque l’on est assuré *a priori* que certains nombres réels échapperont toujours au calcul[131]. De ce point de vue, Turing ne fait que suivre une tradition courante depuis Cantor et que l’on retrouve ensuite chez Richard[132]. Comme Turing fait usage de l’argument de Richard[133] en l’adaptant au problème qui est le sien, il est assez naturel qu’il se place du

---

[130] Cf. **G. Cantor**, “Sur une question élémentaire de la théorie des multiplicités” (1892), Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 1, pp. 75-78, traduction française dans [*Logique et Fondement des mathématiques*, Anthologie (1850-1914), **F. Rivenc** et **P. de Rouilhan dirs.**, op. cit.], pp. 200-203.

[131] C’est pourquoi R. Gandy peut déclarer : «Turing considère les “nombres (réels) calculables”; en fait, ce qu’il considère sont les fonctions calculables totales avec argument positif entier ayant pour valeur 0 ou 1. Il commence son analyse détaillée (dans “On Computable Numbers…” pp. 249-258) par dire : “L’enjeu véritable est la question “Quels sont les processus possibles qui peuvent être effectués dans le calcul d’un nombre réel ?”. Il y a une différence significative entre cette question et la question (“Qu’est-ce qu’une fonction calculable ?”) que d’autres auteurs se posent. Turing s’est pour ainsi dire placé lui-même dans la bonne direction». **R. O. Gandy**, “The Confluence of Ideas in 1936” in [*The Universal Turing machine*, **R. Herken ed.**, op. cit.], p. 80.

[132] Cf. **Richard J.**, “Les principes des mathématiques et le problème des ensembles”, (1905), Revue générale des sciences pures et appliquées, 16, 541, republié dans [*Logique et Fondement des mathématiques*, Anthologie (1850-1914), **F. Rivenc** et **P. de Rouilhan dirs.**, Payot, Paris, 1992], pp. 271-275.

[133] Notons qu’il n’en fait qu’indirectement usage puisque Turing ne fait pas référence à Richard mais, au § 8 de “On Computable Numbers …”, au livre de Hobson, *Theory of Functions of a Real Variable* (2nd edition, 1921), qui le décrit.

même point de vue que lui, à savoir celui de l'accessibilité à l'ensemble numérique des réels.

Pour réussir à cerner au mieux la notion de calculabilité, il faut donc réussir à produire un résultat d'impossibilité, c'est-à-dire un cas négatif où les limites du calcul sont atteintes. Ce cas, c'est le problème de l'arrêt. Une fois en possession de ce résultat, il est possible de rabattre sur lui la question de l'*Entscheidungsproblem*, si l'on parvient à montrer que la solution positive à l'*Entscheidungsproblem* exigerait de résoudre positivement le problème de l'arrêt. Comme la solution au problème de l'arrêt est négative, on en déduit qu'il faut répondre négativement à la question de l'*Entscheidungsproblem*. Telle est la démarche adoptée par Turing dans "On Computable Numbers …" au dernier paragraphe de l'article (§ 11).

12. Aspects mathématiques de la solution adoptée par Turing

Il faut exposer un certain nombre de définitions qui constituent le point de départ de Turing pour parvenir à décrire les aspects proprement mathématiques de sa démarche.

121. Machines circulaires et machines non-circulaires

Au paragraphe 2 de "On Computable Numbers …", Turing distingue plusieurs types de machines.

La première distinction concerne les "machines-*a*" et les "machines-*c*". Turing n'utilise, par la suite, que le cas des machines-*a* : une "machine-*a*" est une "machine automatique", c'est-à-dire une machine qui est entièrement déterminée par ses configurations. Une machine-*c* en revanche est une "machine à choix" qui n'est que partiellement déterminée par ses configurations.

La deuxième distinction, capitale pour la bonne marche de la démonstration du problème de l'arrêt, est la distinction entre machine circulaire et machine non-circulaire. Une machine est dite circulaire quand elle ne produit en sortie qu'un nombre fini de 0 et de 1. Dans le cas contraire, il s'agit d'une machine non-circulaire.

Seule une machine non-circulaire peut produire en sortie le développement décimal correspondant à un nombre réel, puisque seule une machine de ce type peut calculer la suite infinie des décimales correspondant à un nombre réel. On définit alors une suite calculable comme une suite qui peut être calculée par une machine non-circulaire.

122. La position du problème de l'arrêt

Au vu de la définition de la machine non-circulaire, on doit d'abord dissiper une objection.

On pourrait en effet se demander si cette définition n'est pas arbitraire et s'il existe réellement une machine de ce type : quand on décrit la table d'instructions d'une machine dont le but est de calculer le développement décimal d'une nombre réel, peut-on être sûr *a priori* que ce seront bien les décimales recherchées que la machine va écrire en sortie ? La réponse est “non” mais cette réponse n'implique pas une quelconque infériorité de la notion de machine par rapport à celle d'algorithme : on n'est en effet pas plus certain du résultat quand il s'agit d'un algorithme que lorsqu'il s'agit d'une machine. Nous avions déjà mentionné ce fait[134] quand il s'était agi de décrire la spécificité de l'acte de calcul : on accordait à l'algorithme caractérisant la fonction successeur un statut universel sans évidemment vérifier la validité de la règle pour tous les entiers. Dans les deux cas, algorithme ou machine, la présupposition d'universalité est assumée par la même instance, le sujet connaissant qui est décrit comme intuition. Ainsi, de même que l'intuition *reconnaît* l'existence d'un algorithme sur un nombre réduit d'instances le vérifiant, de même doit-elle *reconnaître* l'existence de machines non-circulaires, dès lors qu'on peut donner leurs tables d'instructions. Aussi l'objection n'a pas de valeur dans le cadre de la thèse de Turing qui pose en principe l'équivalence des notions d'algorithme et de machine de Turing.

En fait, du point de vue de la thèse de Turing qui a toujours pour cadre argumentatif la question : “l'intuition peut-elle trouver une machine qui

---

[134] Cf. supra chapitre II, § 222. 1.

corresponde à l'algorithme … ?" - ou, plus brièvement : "peut-on trouver une machine qui … ?" -, la question que l'on peut poser au sujet des machines non-circulaires apparaît sous une forme plus complexe dans la mesure où elle implique de dédoubler la notion de machine : peut-on trouver une *machine* qui décide si une *machine* donnée produira en sortie la suite infinie des décimales d'un nombre réel, c'est-à-dire si une machine donnée est non-circulaire ? C'est la façon dont "On Computable Numbers …" expose, au paragraphe 8, ce qu'il est maintenant convenu d'appeler le "problème de l'arrêt". On sait, grâce à la démonstration que nous avons exposée au chapitre II[135], que la réponse à ce problème est "non" : il n'y a pas de machine qui pourrait parvenir à cette décision. Mais l'itinéraire emprunté par Turing pour parvenir cette conclusion est très différent de la façon dont on l'a exposé plus haut en suivant les présentations classiques.

Quelle est la spécificité de la démonstration exposée par Turing ? Celle-ci se distingue de la démonstration devenue classique sur deux points essentiels : elle fait usage de la notion de machine de Turing universelle ainsi que de ce qu'il est convenu d'appeler l'argument de diagonalisation. Ces deux points confèrent à la démonstration originelle de Turing un aspect plus mathématique que logique. Tentons de les décrire.

La question que l'on vient de poser dans le cadre de la thèse de Turing ("une machine étant donnée, peut-on trouver une machine qui décide si la première produira en sortie la suite infinie des décimales d'un nombre réel ?") est une question particulière dans la mesure où elle porte sur une machine particulière. Elle est immédiatement généralisable en la question suivante : "peut-on trouver une machine qui, prenant une autre machine en entrée, décide si cette dernière produira en sortie la suite infinie d'un nombre réel ?".

Pour réussir à répondre à la question de façon précise, il faut donc avoir les moyens de constituer une liste de ces machines et pour ce faire, deux préalables sont requis : d'une part, il faut leur attribuer à chacune un numéro et d'autre part il faut avoir les moyens de passer cette liste en revue de façon mécanique. Étudions ces deux conditions.

---

135 Cf. supra chapitre II, § 231.

122. 1. La liste des suites calculables

Pour ce qui est du premier point, on doit commencer par donner un numéro de description à chaque machine calculant une suite infinie. Si l'on reprend l'exemple que nous avions donné au chapitre II § 215 d'une machine qui calcule la suite infinie 0101010101 … et qui correspond en fait au premier exemple donné par Turing dans "On Computable Numbers …"[136], le numéro d'une telle machine dans la liste des machines est attribué de la façon suivante.

Comme on l'a vu au § 216 du chapitre II, la table d'instructions correspondant à une machine permet de donner une description complète de la structure logique de la procédure qu'elle exécute. Cette table est composée d'un ensemble de quintuplets de la forme : (1) ancien état q_i; (2) symbole lu s_j; (3) nouvel état q_{ij}; (4) symbole écrit s_{ij}; (5) direction du mouvement d_{ij}. Chaque élément de quintuplets de la table d'instructions peut être codé par une lettre suivie d'une autre répétée autant de fois que l'indice de la lettre qu'elle code : par exemple, un état q_i est codé par une lettre suivie d'une seconde répétée i fois. Chaque lettre peut ensuite être codée sous la forme d'un numéral. Le nombre représenté par l'ensemble formé de ces numéraux est appelé Nombre Descriptif de la machine.

Par exemple, dans le cas de la table d'instructions correspondant à la machine qui calcule la suite infinie 01010101 … on doit coder quatre quintuplets, puisque la machine est susceptible d'entrer dans quatre états.

Si l'on appelle les quatre états q_1, q_2, q_3 et q_4; l'absence de symbole et les deux symboles 0 et 1, S_0, S_1, et S_2 et si à "à droite" est appelé "D", alors on a les quatre quintuplets suivants : $q_1S_0S_1Dq_2$; $q_2S_0S_0Dq_3$; $q_3S_0S_2Dq_4$; $q_4S_0S_0Dq_1$.

On obtient un code sous forme de lettres de la manière suivante :

q_i est remplacé par la lettre "E" suivie de la lettre "A" répétée i fois.

S_j est remplacé par la lettre "E" suivie de la lettre "C" répétée j fois.

[136] **A. M. Turing**, (1936), "On Computable Numbers (…)" § 3, dans [**M. Davis ed.**, *The Undecidable*, op. cit.], p. 119.

“D” est conservé tel que, de même que “;”.

Ce qui donne pour le premier quintuplet : EAEECREAA.

On peut ensuite coder cette expression par des numéraux :

“A”	est remplacé par	1
“C”	est remplacé par	2
“E”	est remplacé par	3
“R”	est remplacé par	5
“;”	est remplacé par	7

Les éléments du premier quintuplet sont alors codables sous la forme : 3133253117.

En exécutant la même traduction pour les trois quintuplets restants, on obtient alors comme Nombre Descriptif le nombre suivant :

3133253117311335311173111332253111173111133531731323253117

Ce nombre décrit la machine qui calcule la suite 010101 … et elle seule. D’autres nombres pourraient décrire la même machine (par exemple, si l’on rajoutait des états non nécessaires à l’effectuation du calcul) mais ce nombre ne décrit que la machine qui calcule la suite 010101 … .

Il est possible de généraliser ce point de vue et de coder toute table d’instructions de machine de Turing calculant une suite. C’est pourquoi Turing peut conclure[137] :

> **«A chaque suite calculable correspond au moins un nombre descriptif, alors qu’à aucun nombre descriptif ne correspond plus d’une suite calculable. Les suites et les nombres calculables sont donc énumérables».**

Le second point qu’il faut maintenant étudier est celui des moyens dont on dispose pour passer en revue la liste en question.

122. 2. Constitution mécanique de la liste des suites calculables

La constitution de la liste des suites calculables doit elle-même relever du mécanique. Cette condition implique en fait une impossibilité : il ne peut pas y avoir une énumération mécanique des suites calculables. Turing use pour ce faire

---

[137] **A. M. Turing**, (1936), “On Computable Numbers (…)” § 5, dans [**M. Davis ed.**, *The Undecidable*, op. cit.], p. 127.

de l’argument de diagonalisation.

122. 21. Argument de diagonalisation

Turing rejette tout d’abord, au début du paragraphe 8, un usage inadéquat de l’argument de diagonalisation : en partant de l’hypothèse que l’on peut énumérer les suites calculables (quels que soient les moyens d’effectuer cette énumération), on peut construire à partir de la liste une suite qui semble ne pas appartenir à l’énumération tout en ayant toutes les apparences de la calculabilité : la suite obtenue par “diagonalisation”[138]. Mais ce raisonnement, comme Turing le fait remarquer immédiatement, implique de croire au départ qu’il est possible d’effectuer la mise en liste des suites par des moyens finis; or c’est précisément ce qui n’est pas possible. Il faut donc redresser l’argument pour lui accorder une portée.

L’argument devient alors : s’il était possible de dresser la liste des machines non-circulaires, on pourrait calculer la suite β, que l’on obtient par diagonalisation. Donc la suite β serait à la fois calculable et incalculable; il n’existe donc pas de procédure générale permettant d’énumérer la liste des machines non-circulaires. Mais comme Turing l’indique au paragraphe 6, il ne retient pas cette façon de procéder, qu’il juge pourtant être «la preuve la plus simple et la plus directe» et ce, parce qu’elle pourrait laisser au lecteur l’impression qu’«il doit y avoir quelque chose qui ne va pas». Turing ne dit pas expressément pourquoi le lecteur pourrait avoir ce sentiment. On remarque seulement que le “détour” emprunté par Turing dans la démonstration qu’il propose implique une argumentation dans laquelle entre en jeu la notion de machine universelle. On peut en inférer[139], que c’est parce que la démonstration

---

[138] Cf. **A. M. Turing**, (1936), “On Computable Numbers (…)” § 8, dans [**M. Davis ed.**, *The Undecidable*, op. cit.], p. 132 : «Ou nous pourrions appliquer le processus diagonal. “Si les suites calculables sont énumérables, posons que a_n est la n-ième suite calculable et que $\Phi_n(m)$ est le m-ième chiffre dans a_n. Soit β la suite qui a pour n-ième chiffre $1 - \Phi_n(n)$. Puisque β est calculable, il existe un nombre K tel que $1 - \Phi_n(n) = \Phi_K(n)$ pour tout n. Pour $n = K$, nous obtenons $1 = 2\Phi_n(K)$, c’est-à-dire que 1 est pair. C’est impossible. Donc les suites calculables ne sont pas énumérables”».

[139] Cf. **J. Mosconi**, *La constitution de la théorie des automates,* op. cit., p. 45.

ne se laisse pas facilement interpréter en termes mécaniques que Turing a préféré en présenter une version légèrement différente qui s'intègre plus facilement au cadre de la thèse de Turing. Celui-ci fait d'ailleurs lui-même remarquer que la démonstration qu'il va exposer permet de mettre en lumière la notion de machine en précisant ce que l'on doit entendre par machine "non-circulaire".

La démonstration qu'il expose entre dans la catégorie du paradoxe de la liste ouverte qui devrait se précéder elle-même. L'idée même d'une liste infinie des machines non-circulaires finit par se révéler contradictoire : d'une part cette liste doit être produite par une machine non-circulaire mais d'autre part la machine non-circulaire produisant la liste ne peut être située nulle part dans la liste elle-même. Si en effet elle était située quelque part, la liste serait en fait produite par une machine circulaire : la liste serait en effet close sur elle-même puisqu'elle contiendrait une machine clôturant la liste. Une machine ne pouvant être à la fois circulaire et non-circulaire, il est contradictoire de vouloir dresser la liste des machines non-circulaires. Il y a donc des machines non-circulaires inaccessibles mécaniquement.

Le nerf de l'argument repose sur l'existence d'une machine décisionnelle "D" capable de décider si une machine donnée est ou non circulaire, c'est-à-dire capable de dresser une ligne de partage entre deux listes, la liste des machines non-circulaires et la liste des machines circulaires. Une machine "M" formée d'une machine universelle "U" combinée avec "D", peut dès lors imiter le calcul de n'importe quelle machine non-circulaire rangée dans la liste en calculant n'importe quelle suite jusqu'à n'importe quel rang. Quand M en vient à calculer une suite β' qui occupe un rang n dans la liste, apparaît une difficulté de rangement entre le rang n de la liste et le rang n de la suite : pour calculer le rang n de la suite, la machine M devrait avoir rangé la suite dans la liste au rang n en ayant déjà calculé n.

Ainsi, dans l'argument tel qu'il est présenté par Turing, la machine universelle sert-elle seulement à imiter le calcul d'autres machines, selon la fonction qui est toujours la sienne, mais ne participe-t-elle pas à la constitution supposée de la liste des machines non-circulaires, fonction dévolue à la pseudomachine "D". On voit donc que la preuve de l'insolubilité du problème de l'arrêt ne doit que peu de chose à la notion de machine universelle. Celle-ci n'apporte finalement qu'une facilité technique dans l'imitation du calcul des machines mais sans effectuer le partage "décisionnel" entre la liste des machines circulaires et la liste des machines non-circulaires. Comme nous l'avions déjà remarqué au chapitre II, l'utilisation de la machine universelle est seconde par rapport à la présupposition première, à savoir qu'il existe une machine (dont on ne précise pas le type) susceptible de répondre à la question de la nature globale de toute suite donnée, c'est-à-dire à la question : "est-ce vraiment une suite infinie ou pas ?" ou, traduite selon la thèse de Turing : "est ce une machine non-circulaire ou pas ?". La machine universelle permet seulement de calculer n'importe quelle place particulière de n'importe quelle suite particulière. Aucune machine, *qu'elle soit ou non universelle*, ne peut apprécier le caractère infini d'une suite - pas plus, d'ailleurs, que l'être humain. Dès lors, pourquoi Turing invoque-t-il la notion de machine universelle dont l'usage reste quelque peu périphérique puisqu'il ne permet pas de faire avancer de façon décisive la solution au problème de l'arrêt ?

122. 22. Justification psychologique à l'utilisation de la notion de machine universelle

Il faut, pour le comprendre, revenir un instant sur le processus d'intuition tel que nous avons déjà eu l'occasion de le décrire.

La reconnaissance de la présence d'une suite consiste pour l'intuition à reconnaître la présence d'une *identité* dans le développement d'une énumération, identité qui confère à l'énumération le statut de suite : par exemple, on reconnaît une identité entre les couples de chiffres dans l'énumération 0101010101 … . Aussi l'aspect infini de la suite ne doit-il pas être considéré comme négatif parce qu'indéterminé : il est au contraire la manifestation de la présence de l'identité, quel que soit le nombre d'éléments considéré, qui devient une simple question empirique. Du point de vue de la thèse de Turing, la notion de machine serait la traduction en termes mécaniques de l'aspect identitaire de cette reconnaissance. En effet, l'existence de la notion de machine montre l'aspect foncièrement identique de tout processus de calcul puisque, quel que soit le calcul envisagé, il est toujours effectuable par une machine.

Mais ce n'est que le premier aspect du processus intuitif. En effet, ce qui est remarquable, c'est que non seulement l'intuition reconnaisse l'identité d'une suite dans ce qui n'est qu'énumération sans ordre, mais encore qu'elle reconnaisse dans chacune de ces reconnaissances la présence identique d'un même acte : *le don d'identité* lui-même. Cette capacité trouve, bien que Turing ne le mentionne pas, son exacte réplique au sein de la thèse de Turing : non seulement on peut mécaniser la constitution des suites, c'est-à-dire la reconnaissance de l'identité dans les énumérations, mais on peut mécaniser cet acte de constitution, à savoir le don d'identité lui-même : c'est précisément le rôle imparti à la machine universelle.

Ainsi l'utilisation de la notion de machine universelle par Turing dans le fil de son argumentation touchant le problème de l'arrêt ne vise-t-elle pas à faire progresser l'argumentation elle-même. En revanche, l'existence de la notion de machine universelle a un intérêt psychologique[140] dans la mesure où elle montre la vraie nature de l'intuition dans le cadre de la thèse de Turing : la machine universelle n'est pas seulement la traduction en termes mécaniques d'une notion qui pourrait être pensée par le biais de la notion d'algorithme mais elle est l'expression mécanique d'un *acte* de pensée. Plus précisément, la machine universelle *est* la pensée en acte en tant que celle-ci se manifeste à travers le schème du calcul. C'est pourquoi Turing ne se réfère pas à la thèse - et à la possibilité qu'elle institue de traduction en termes mécaniques de la notion d'algorithme - pour justifier l'existence de la machine universelle : au paragraphe 7, il présente d'emblée la table d'instructions d'une machine de ce type.

C'est en ce sens que la notion est capitale pour la constitution d'une psychologie : ce n'est pas à un niveau seulement linguistique de traductibilité qu'elle apparaît mais à un niveau réel, celui de l'acte de pensée lui-même, comme mise au jour d'un *schème*.

---

[140] Nous l'avions déjà noté au chapitre 2, § 221, quand nous avions remarqué que l'apparition de la notion de machine universelle de Turing ne se justifiait pas dans l'argumentation de Turing si ce n'est pour apporter une corroboration psychologique à la "thèse de Turing".

Une question demeure ouverte : comment s'opère par l'intuition la reconnaissance de la présence de l'identité dans ce qui n'est au départ qu'énumération sans ordre[141] ? Dans le cadre de la thèse de Turing, cette question s'énonce sous la forme suivante : comment s'opère par l'intuition la reconnaissance de l'existence de la machine universelle ? Autrement dit, comment s'opère par la pensée la mise au jour du schème mécanique de la pensée ? A cette dernière question, Turing répond de façon presque désinvolte au paragraphe 6 de "On Computable Numbers …", quand il mentionne l'existence de la machine universelle pour la première fois :

> **«Il est possible d'inventer une machine unique qui peut être utilisée pour calculer n'importe quelle suite calculable. Si cette machine U est munie d'un ruban au début duquel est inscrite la description standard d'une machine à calculer M, alors U calculera la même suite que M».**

Mais c'est l'accès psychologique radicalement nouveau à cette possibilité, ce que Turing appelle "l'invention d'une machine unique", qui reste sans justification. D'où provient cette invention ? La question est fondamentale dans une théorie de la psychologie puisqu'il est nécessaire dans ce cas de rendre raison de l'accès psychologique aux schèmes de pensée. Notons seulement, pour l'instant, que Turing n'aborde la question que par prétérition, jugeant sans doute qu'un article de mathématiques n'est pas le lieu adéquat pour en débattre[142].

Après "On Computable Numbers …" néanmoins, Turing revient, dans une perspective logique, sur la question de la détermination psychologique du schème mécanique de la pensée. C'est ce qu'il nous faut étudier maintenant.

2. Le point de vue logique adopté par Turing

C'est en effet en abordant des questions d'ordre strictement logique que le problème de la nature de la psychologie réapparaît dans l'œuvre scientifique de

---

[141] Le fonctionnement cognitif de cette reconnaissance ne nous intéresse pas ici, bien qu'il soit à la fois capital et mystérieux. On tentera d'apporter des éléments permettant d'en rendre compte dans le cours de la deuxième partie.

[142] Ce ne sera plus le cas en 1950 quand Turing écrit "Computing Machinery and Intelligence", comme nous le verrons dans la deuxième partie.

Turing. Remarquons toutefois que, dès “On Computable Numbers …”, Turing a adopté une perspective strictement logique en vue d’appliquer le résultat du problème de l’arrêt à l’*Entscheidungsproblem* sans apporter de modifications à sa conception de la psychologie. Turing expose ainsi cette “application” :

> **«Je propose donc de montrer qu’il ne peut y avoir de procédé général pour déterminer si une formule donnée du calcul fonctionnel K est démontrable, c’est-à-dire qu’il ne peut pas y avoir de machine à qui l’on aurait fourni une de ces formules quelconques U et qui finirait par dire si U est ou non démontrable».**

Du point de vue de la théorie de la psychologie, l’application du problème de l’arrêt au cas de l’Entscheidungsproblem n’apporte rien de nouveau si ce n’est que la notion de machine universelle n’y apparaît pas. Ainsi Turing montre-t-il directement en quelque sorte, que le cas de la démontrabilité des formules du calcul des prédicats exigerait de posséder cette machine “décisionnelle” dont on a montré qu’elle faisait nécessairement défaut.

En revanche, le cas de “Systems of Logic based on Ordinals” apporte, dans un contexte proprement logique, des précisions sur le contenu d’une théorie de la psychologie, parce que Turing essaye d’y préciser ce qu’il faut entendre par faculté d’intuition. On avait décrit l’intuition comme la faculté capable de reconnaître une identité dans ce qui se donne au départ comme sans ordre. On avait remarqué à ce propos que la différence entre le fini et l’infini, pour ce qui est du nombre des éléments pris en considération, était une simple question empirique. C’est cette idée qui se trouve exploitée dans “Systems of Logic based on Ordinals”, dans lequel Turing montre comment articuler procédure mécanique et récurrence transfinie :

> **«Le célèbre théorème de Gödel 1931 montre que, d’un certain point de vue, tout système de logique est incomplet, mais il indique en même temps les moyens par lesquels on peut obtenir à partir d’un système L de logique un système L' plus complet. En répétant le processus, nous obtenons une suite L, $L_1 = L'$, $L_2 = L'_1$, … chacun plus complet que le précédent. Une logique L_ω peut être alors construite dans laquelle les théorèmes démontrables sont la totalité des théorèmes démontrables à l’aide des logiques L, L_1, L_2, … En procédant de cette façon, nous pouvons associer un système de logique avec un ordinal constructif quelconque. On peut se demander si une suite de logique de ce type est complète au sens où à un problème A quelconque correspond un ordinal α tel que A est démontrable au moyen de la logique L_α».**

Le fait que l’on puisse étudier une succession de systèmes axiomatiques indexés sur certains ordinaux transfinis est encore une manifestation du pouvoir propre à l'intuition : à l'aspect proprement mécanique lié à l'application des règles et des axiomes, on doit ajouter la possibilité d'opérer une récurrence transfinie en reconnaissant dans un ordinal une notation susceptible de pouvoir représenter une logique donnée. A titre d'exemple, Turing aborde au paragraphe 10 la question d'un équivalent à l'hypothèse du continu de Cantor, selon laquelle il y a une correspondance bi-univoque entre l'ensemble P(ω) des sous-ensembles de ω et l'ensemble de tous les ordinaux plus petits que le premier ordinal non-dénombrable ω_1. Turing prend comme équivalent à l'ensemble P(ω) l'ensemble des suites calculables de 0 et de 1 (correspondant aux réels calculables) ou l'ensemble des Nombres Descriptifs des machines qui calculent les suites calculables, tout en faisant remarquer que son choix est arbitraire. Comme équivalent à ω_1, il prend le plus petit ordinal non-constructible ω_1^{CK}. Turing montre qu'on ne peut pas établir de correspondance bi-univoque entre ces ensembles parce qu'on ne peut pas trouver de fonction calculable qui calculerait la correspondance en question. Il ne s'agit, pour Turing, que d'un exemple puisque, comme il le fait remarquer lui-même, ce résultat «n'a pas d'intérêt réel pour ce qui est de l'hypothèse du continu classique».

On remarque qu'en essayant de déterminer ce qu'il faut entendre par intuition, Turing ne tente pas de sortir du cadre théorique qui est le sien et qui est celui de la thèse de Turing. Plus précisément, bien qu'il reconnaisse que la faculté d'intuition ne peut pas être définie par le biais de la notion de machine, *il ne cherche pas à la situer au-delà du mécanique*. Aussi, même lorsque Turing tente de concevoir en quel sens il serait possible de contourner les limitations internes des axiomatiques en utilisant la possibilité d'engendrer des systèmes non-dénombrables d'axiomes, est-ce bien par rapport au calcul que la notion d'intuition peut prendre un sens déterminé[143]. En ce sens, l'idée même d'une

---

[143] Remarquons de ce point de vue que Gödel faisait peut-être un mauvais procès à Turing quand il l'accusait d'avoir commis «une erreur philosophique» dans "On Computable Numbers …" § 9, consistant à ne pas avoir suffisamment étudié la possibilité d'une convergence transfinie d'axiomatiques limitées, puisque cette possibilité définit en partie le projet de "Systems of Logic based on Ordinals". Cf. **K. Gödel**, "Some Remarks on the Undecidability Results", 1972a,

faculté d'intuition du non-dénombrable ne vise absolument pas à mettre à mal la théorie de la calculabilité élaborée antérieurement par Turing parce que c'est toujours le même ressort identitaire qui est utilisé dans le cas des logiques ordinales. L'utilisation du transfini a au contraire pour effet de transformer cette théorie en faisant évoluer le concept même de calculabilité qui, de notion *absolue* devient *relative* au système axiomatique envisagé[144]. Dès lors l'intuition, toute non-mécanique qu'elle soit dans l'absolu, se manifeste-t-elle cependant dans la sphère du calcul qu'elle ne transcende relativement que par degrés. Aussi n'a-t-elle de sens que dans le rapport qu'elle entretient à la sphère du calcul, seule habilitée à représenter ce que l'on entend précisément par procédure bien définie.

21. Décision et machine à oracle

Contrairement à la démarche de "On Computable Numbers …", qui tendait à assimiler intuition humaine et fonctionnement mécanique, Turing adopte donc dans "Systems of Logic based on Ordinals" un point de vue qui les distingue sans les opposer, car si l'intuition n'est pas mécanique, c'est cependant au sein de la sphère du mécanique qu'elle peut se manifester.

Pour ce faire, Turing se dote, au paragraphe 4, d'une capacité psychologique particulière, qu'il appelle un "oracle" et qui est capable d'opérer une décision du même type que celle que dont on aurait besoin pour résoudre le problème de l'arrêt. Turing montre ensuite qu'il est possible de définir à partir de cet "oracle" une machine d'un nouveau type, qu'il appelle "machine à oracle" (machine-*o*) :

> **«Supposons que l'on nous fournisse un moyen non-spécifié de résoudre des problèmes numériques; une sorte d'oracle pour ainsi dire. Nous ne nous étendrons pas davantage**

---

reproduit dans **K. Gödel**, [*Collected Works*, t. II, Oxford University Press, Oxford, 1990], p. 306. J. Webb fait une remarque du même type dans son introduction au texte de Gödel, ibidem, p. 298, note w : «(…) l'hypothèse de finitude de Turing est parfaitement compatible avec un point de vue "dynamique" sur l'esprit».

144 C'est ce que fait remarquer S. Feferman dans **S. Feferman**, "Turing in the Land of O(z)", dans [*The Universal Turing Machine*, **R. Herken ed.**, Oxford University Press, 1988], p. 127 : « Là [dans cet article] Turing s'attaqua de façon tout à fait systématique à l'idée naturelle consistant à essayer de dépasser l'incomplétude gödelienne des systèmes formels au moyen de principes d'itération transfinie qui serviraient à dépasser localement l'incomplétude».

sur la nature de cet oracle, sauf pour dire qu'il ne peut pas être une machine. En s'aidant de cet oracle, nous pourrions former un nouveau type de machine (que nous appellerons machine-*o*), dont un des processus fondamentaux serait de résoudre un certain problème numérique donné».

Ce type de machine joue le rôle que jouait la machine décisionnelle "D" dans la démonstration du problème de l'arrêt de "On Computable Numbers …". Turing expose ensuite le fonctionnement de ce type de machine en décrivant ce que serait sa table d'instructions :

«Plus précisément, ces machines doivent se comporter ainsi. Les mouvements de la machine sont comme d'habitude déterminés par une table sauf dans le cas où les mouvements sont dans une certaine configuration *o*. Si la machine est dans une certaine configuration interne *o* et si la suite de symboles marquée de l est alors une formule bien formée A, alors la machine se place dans une configuration interne *p* ou *t* selon qu'il est vrai ou faux que A est duale [c'est-à-dire selon que A est vraie ou que ¬ A est vraie]. La décision de savoir ce qui est le cas est référée à l'oracle».

Turing démontre alors, en se référant expressément au paragraphe 8 de "On Computable Numbers …", qu'il est impossible de déterminer si une machine-*o* qui aurait reçue la table d'instructions d'une autre machine-*o* serait capable de décider si la machine-*o* examinée est une machine-*o* non circulaire.

«[…] il n'est pas possible de construire une machine-*o* qui, ayant reçue la description de tout autre machine-*o* pourra déterminer si cette machine est non-circulaire ou pas».

Comment concevoir le rapport entre l'intuition, l'oracle et la notion de machine ? Du point de vue de la thèse de Turing, il faut étudier le rapport entre des facultés psychologiques et leurs expressions mécaniques. Commençons par étudier les expressions mécaniques.

22. Machine-*a*, machine-*c*, machine-*u* et machine-*o*

De "On Computable Numbers …" à "Systems of Logic based on Ordinals", Turing a élaboré quatre types de machines. Décrivons-les.

Une machine-*a* est une "machine automatique" : comme l'indiquait Turing dans "On Computable Numbers …", elle était définie comme une machine entièrement déterminée par ses configurations

Une machine-*c* est une "machine à choix" qui n'est que partiellement déterminée par ses configurations et que Turing, dans “On Computable Numbers …”, opposait à la machine-*a*. Ce type de machine ne nous intéresse pas directement ici.

Une machine-*u* est une machine-*a* qui a la particularité d’être universelle. Cette expression n’apparaît pas directement sous la plume de Turing, bien que dans “On Computable Numbers …”, la machine universelle qu’il utilise soit nommée “U”.

Une machine-*o* est une “machine à oracle” qui est capable de décider de questions numériques quelle qu’elles soient, sans que soient précisées sa nature ni table de configurations.

Trois de ces types nous intéressent ici : les machines-*a*, les machines-*u* et les machine-*o*. Comme nous l’avons déjà remarqué, la différence entre ces types de machines n’est pas une opposition parce que ce sont en fait trois types d’*expression* provenant d’une même source productrice.

De ce point de vue, les expressions provenant de cette source sont *toujours* des machines : aussi la notion de machine est-elle bien un schème général de pensée et “On Computable Numbers …” en est la description mathématique précise. Plus précisément, le schème mécanique s’exprime dans toute sa généralité dans la notion de machine universelle : toute machine-*a* est une expression mécanique qui peut être *imitée* au moyen d’une machine universelle, c’est-à-dire qu’il est toujours possible d’en rattacher l’origine à un acte identique de pensée, c’est-à-dire à son schème. La machine-*u* exprime donc, comme nous l’avions déjà remarqué, un schème général de toute pensée calculante. C’est donc par rapport à ce schème que l’on doit étudier la nature des deux autres types de machines.

Ce schème de pensée peut s’exprimer de deux manières différentes selon qu’il est ou non contradictoire dans le contexte dans lequel il apparaît : soit la schématisation parvient à s’achever et s’exprime sous la forme d’une machine-*a* qui peut être imitée par une machine universelle; soit la schématisation se constitue en machine-*o* et elle est renvoyée dans l’oracle quand le problème numérique étudié, posé en termes mécaniques, se trouve être insoluble, c’est-à-dire ne peut pas achever d’être schématisé. Dans ce cas, la machine-*o* apparaît comme une expression mécanique qui a perdu la trace de son schème.

Il nous faut maintenant décrire le rapport de ces expressions à la source psychologique qui les produit. Cette description constitue le point de vue psychologique adopté par Turing.

3. Le point de vue psychologique adopté par Turing

Nous allons tenter de dégager la façon dont Turing décrit d'un point de vue général l'intervention du psychologique dans le raisonnement mathématique en nous appuyant sur les deux articles de 1936 et de 1939 déjà cités, ainsi que sur un texte de Gödel portant sur la notion de machine de Turing universelle.

31. Conjecture sur le rôle psychologique de la machine universelle

La solution négative au problème de l'arrêt a une conséquence importante sur le statut de la machine universelle : il ne peut pas y avoir de procédure de décision permettant de prédire le comportement d'une machine universelle. Dans la mesure où celle-ci ne fait qu'imiter toutes les machines, qu'elles soient circulaires ou non-circulaires, les états d'une machine universelle sont descriptibles sans que son comportement le soit. Gödel faisait remarquer à ce propos [145]:

> **«Là [dans le cas de la machine de Turing universelle] on pourrait dire que la description complète de son comportement est infini parce que, au vu du fait qu'il n'existe pas une procédure de décision prédisant son comportement, la description complète ne pourrait être donnée que par l'énumération de toutes ces instances. Évidemment, ceci présuppose que seules les descriptions décidables sont considérées comme complètes, mais cela va dans le sens du mode de pensée finitiste. La machine de Turing universelle, dans laquelle le rapport des deux complexités est l'infinité, devrait donc être considérée comme un cas limite des autres mécanismes finis».**

La remarque de Gödel permet de dire qu'il y a *autre chose* dans le schème du calcul que l'imitation, puisque ce concept ne permet pas d'assurer la pleine mise en lumière de la notion de calcul. Qu'y-a-t-il de plus que l'imitation ? Seulement le fait qu'il y a autre chose dans l'acte de calcul que sa pure et simple

---

[145] Lettre à Arthur Burks, cité dans l'introduction à **J. Von Neumann**, *Theory of Self-Reproducing Automata*, ed. A. Burks, University of Illinois Press, Urbana and London, 1966, p. 56.

traduction symbolique par le biais d'un codage. Psychologiquement, cette remarque a une importance considérable puisqu'elle laisse une place pour l'existence d'une capacité qui n'est pas d'emblée linguistique - au sens où elle permettrait seulement une traduction par codage - mais bien psychologique. De plus, cette faculté ne se situe pas dans un "ailleurs" du mécanique mais se laisse décrire par l'aspect non-prédictible de la machine universelle. C'est sans doute la raison pour laquelle la notion de machine universelle apparaissait à Turing comme la notion qui rend psychologiquement le mieux compte de ce qu'il faut entendre par calcul. Turing va préciser ultérieurement cet aspect grâce à la notion d'oracle.

32. Intuition et oracle

Au paragraphe 11 de "Systems of Logic based on Ordinals"[146], Turing décrit cette source en exposant de façon très générale ce qu'il entend par raisonnement mathématique :

> **«Le raisonnement mathématique peut être considéré de façon schématique comme l'exercice d'une combinaison de facultés que nous pouvons appeler l'intuition et l'ingéniosité. L'activité de l'intuition consiste à produire des jugements spontanés qui ne sont pas le résultat de chaînes conscientes de raisonnement. Ces jugements sont souvent mais en aucune façon invariablement corrects (en laissant de côté ce qu'il faut entendre par "correct"). […]. L'exercice de l'ingéniosité en mathématique consiste à aider l'intuition par des arrangements adéquats de propositions et peut-être par des figures géométriques ou des dessins. […]. Les rôles joués par ces deux facultés diffèrent évidemment selon les occasions et selon les mathématiciens. Cet aspect arbitraire peut être supprimé en introduisant une logique formelle. Dans les temps pré-gödeliens, certains pensaient que […] la nécessité d'un recours à l'intuition pourrait être entièrement éliminé. […] Nous avons essayé de voir jusqu'où il était possible d'éliminer l'intuition. Nous ne nous préoccupons pas de savoir quelle quantité d'ingéniosité est requise et nous faisons donc l'hypothèse qu'elle est disponible en quantité illimitée».**

Ce texte ne mentionne pas la notion d'oracle et décrit seulement le rapport entre la faculté d'intuition et ce que nous avons appelé le schème mécanique appelé par Turing "ingéniosité". On peut néanmoins à partir de lui réussir à préciser la place qu'occupent respectivement le schème mécanique, l'intuition et l'oracle.

Il est remarquable de constater que le jeu entre ces deux facultés, intuition

---

[146] Cf. **A. M. Turing**, (1939), "Systems of Logic based on Ordinals" Proceedings of the London Mathematical Society, ser. 2, vol. 45 : 161-228, § 4 reprint in [**M. Davis ed.**, *The Undecidable* op. cit.], pp. 208-209.

et schème mécanique, est référé par Turing à une activité *unique*. Le partage qu'instaure Turing entre l'intuition et le schème mécanique n'est donc pas le partage entre deux activités mais entre deux facultés au sein d'une activité mathématique unique. Dès lors, l'intuition n'est pas une faculté qui serait située dans un au-delà de l'effectif à tout jamais inaccessible : elle est seulement inaccessible de façon parcellaire au sein de l'activité mathématique. Comme le fait remarquer Turing, le partage se situe donc au sein de l'activité de calcul, interprété comme l'activité mathématique de pensée en général, *entre le non-conscient et le conscient* et non entre deux facultés conscientes dont l'une (le schème mécanique) aurait un fonctionnement mathématiquement descriptible et l'autre (l'intuition) aurait un fonctionnement cognitif mystérieux[147]. La différence entre le non-conscient et le conscient permet donc de caractériser de façon unique l'activité mathématique et de situer le rapport des facultés en elle dans une perspective dynamique visant à *rendre conscient* ce qui ne l'est pas, tout en sachant que cette tâche n'est jamais achevée.

Quels sont les rapports qu'entretiennent alors la faculté d'intuition et la faculté d'oracle ?

On pourrait caractériser l'oracle comme la face cachée de l'intuition. Il ne faut donc pas les opposer parce qu'elles caractérisent en fait la même faculté psychologique à des niveaux de conscientisation différents. Ces niveaux de conscientisation sont aussi, comme nous venons de le voir, des niveaux de schématisation : quand la source originelle de l'activité mathématique parvient à schématiser complètement son expression sous la forme d'une machine-*a*, elle se manifeste comme intuition, c'est-à-dire comme faculté accessible partiellement par le biais du schème, tandis que lorsque cette même source originelle ne parvient pas à se schématiser complètement et s'exprime sous la forme d'une machine-*o*, elle fait retour vers sa source qui ne se manifeste alors que comme pure activité expressive (l'oracle). L'oracle est donc seulement le processus par lequel une expression donnée en termes mécaniques (la machine-*o*) est

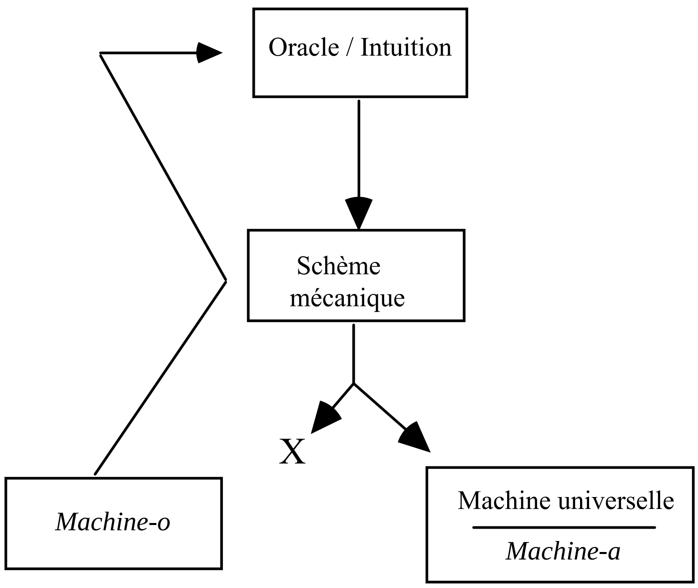
---

[147] Cf. sur ce débat, **P. Maddy**, *Realism in Mathematics*, Clarendon Press, Oxford, 1990 où la question du fonctionnement cognitif d'une intuition du non-dénombrable est discutée et finit par aboutir à une aporie.

mécaniquement renvoyée à sa source psychologique sans passer par le biais du schème mécanique de pensée.

De ce point de vue, la source psychologique des deux types d'expression (machine-*a* et machine-*o*) peut-être décrite soit comme manifestation partielle mais stabilisée (l'intuition), soit comme renvoi direct à la faculté psychologique productrice, qui en devient inaccessible (l'oracle).

On peut alors représenter les rapports de l'intuition et de l'oracle par le schéma suivant :



Pour tenter d'adopter une perspective globale sur la façon dont la psychologie apparaît dans les textes mathématiques et logiques de Turing, il faut revenir à ce qui en constitue le fondement, à savoir la thèse de Turing et comparer celle-ci avec d'autres thèses portant sur le calcul.

33. Les thèses de Turing

On trouve habituellement dans la littérature l'expression de “Thèse de Church-Turing” pour désigner l'énoncé qui assigne à la notion intuitive d'algorithme une "traduction" en termes formels. Mais, comme nous allons le voir, malgré l'équivalence formelle des différentes thèses sur la calculabilité, les présentations diffèrent par leur contenu. Ce sont ces différences qu'il faut essayer maintenant de dégager.

331. La thèse de Church

Cette thèse concernant la nature de la calculabilité a une signification directement mathématique parce qu'elle vise la définition d'une classe de fonctions. La "thèse de Church" comme on a pris l'habitude de l'appeler en suivant Kleene, fut énoncée pour la première fois en 1934 puis sous une forme plus générale en 1936[148]:

> **«Nous définissons maintenant la notion, déjà discutée, de fonction effectivement calculable d'entiers positifs en l'identifiant avec la notion de fonction récursive d'entiers positifs[18] (ou de fonction λ-définissable d'entiers positifs). On considère que cette définition est justifiée par les remarques qui suivent, si tant est qu'on puisse fournir une justification positive au choix d'une définition formelle devant correspondre à une notion intuitive.**
>
> **[18] La question de la relation entre la calculabilité effective et la récursivité (à laquelle on propose ici de répondre en identifiant les deux notions) fut soulevée par Gödel dans une conversation avec l'auteur. La question correspondante de la relation entre la calculabilité effective et la λ-définissabilité avait été indépendamment proposée auparavant par l'auteur.»**

On peut donc présenter la thèse de Church comme suit :

<u>Thèse de Church</u> :

> **Ce qui est considéré intuitivement comme calculable est calculable par fonctions récursives ou par fonctions λ–définissables.**

L'accent est mis sur la traductibilité des formulations les unes dans les autres, et c'est en se reposant sur cette inter-traductibilité que Church suppose que

---

[148] Cf. **M. Davis**, "Why Gödel didn't have Church's Thesis", <u>Information and Control</u> 54, (1982) qui cite, p. 8, une lettre de Church à Kleene de 1934 et aussi **A. Church**, "An unsolvable problem of Elementary Number Theory", <u>American Journal of Mathematics</u>, 58, p. 356, reprint dans [**M. Davis ed.**, *The Undecidable*, op. cit.], p. 100 sq. Dans ce texte, la thèse de Church est appelée "Thèse I". Kleene parle explicitement de la "thèse de Church" dans **S. C. Kleene**, *Introduction to Metamathematics*, North-Holland publ. Co., Amsterdam, 1952, p. 317.

la notion de calculabilité est identique quelles que soient les différences dans les présentations. Si cette intertraductibilité est avérée, (ce qui est le cas puisqu'elle est démontrable[149]) il semble qu'il n'y ait plus besoin de se poser des questions sur le caractère nécessairement psychologisant de cette thèse, qui met en rapport une notion intuitive et une notion formelle : on peut considérer la thèse comme une définition conventionnelle, à laquelle il serait toujours possible d'apporter des modifications en temps utile si le besoin s'en faisait sentir, bien que cette éventualité semble fort peu probable, vu la quasi-impossibilité intuitive qu'il y aurait à concevoir un calcul qui serait non-récursif dans le déroulement de ses étapes.

Ce point de vue tend donc à considérer que la *démonstration* de l'identité des différentes présentations de la notion de calculabilité accrédite la thèse et occulte du même coup le caractère *indémontrable* de la thèse en question ainsi que le jeu des facultés psychologiques qui président à sa constitution.

On a vu que tel n'était pas la perspective de Turing, puisque celui-ci explicite le rapport aux facultés psychologiques sous-jacentes. Revenons sur la thèse en question.

332. Retour à la thèse de Turing

La thèse de Turing postulait, comme celle de Church, la traductibilité en termes mécaniques de tout algorithme. Elle avait la forme suivante :

> **Thèse de Turing :**
>
> **Toute fonction calculable par un être humain en suivant un algorithme peut être calculée par une machine de Turing.**

Au vu de la façon dont Turing a présenté le concept de calculabilité par machine de Turing, la thèse peut maintenant s'expliciter en quatre points.

---

[149] Cf. L'idée d'une intertraductibilité, comme le fait remarquer Kreisel, n'a pas été immédiate. Mais une fois qu'elle a été avancée, elle a permis de dégager ce qu'il y avait de commun à toutes les présentations, à savoir l'idéal calculatoire d'une «rigueur informelle». Cf. **G. Kreisel**, "Church's Thesis and the Ideal of Informal Rigour", Notre Dame Journal of Symbolic Logic, 28, 4, October 1987, p. 507.

1. <u>La thèse de Turing contient une analogie</u>

La thèse recouvre en fait une analogie sur quatre termes qui peut s'énoncer sous la forme suivante : la notion d'algorithme est à la notion de machine de Turing ce que la faculté d'intuition est au schème mécanique qui s'exprime dans la notion de machine universelle.

La thèse ne décrit donc pas seulement le rapport de traduction possible entre deux notions, algorithme et machine, mais dégage aussi les conditions psychologiques de la possibilité de ce rapport. Elle met ainsi au jour un schème de pensée, le schème mécanique, qui se retrouve identique à lui-même dans tout acte de la pensée calculante. Aussi la thèse de Turing se trouve-t-elle dédoublée de la manière suivante quand on prête attention aux facultés psychologiques qui lui sont sous-jacentes. On a donc le schéma suivant :

<u>Thèse de Turing n° 1</u>

algorithme de calcul $\longleftrightarrow$ *machine-a*

<u>Thèse de Turing n° 2</u>

Intuition $\longleftrightarrow$ machine-u (schème mécanique

Cette seconde thèse de nature psychologique implique un certain nombre de conséquences touchant la façon de concevoir l'esprit.

2. <u>Les limites de la calculabilité</u>:

L'impossibilité d'une solution positive au problème de l'arrêt a montré que l'application du schème mécanique pouvait rencontrer une résistance. Le domaine du calculable apparaît donc comme ayant des limites. D'un point de vue psychologique, la découverte de ces limites fait prendre conscience de l'existence d'un schème mécanique de pensée. Dès lors, ce schème acquiert une certaine autonomie par rapport à d'autres schèmes possibles et il s'exprime à un niveau général par l'idée d'une machine universelle. Psychologiquement, on a donc tendance à "anthropomorphiser" la notion de machine en lui prêtant une autonomie. On peut énoncer ce fait par le biais d'une analogie : la machine universelle est aux machines particulières ce que le sujet calculant est aux algorithmes. cette analogie repose sur la thèse psychologique suivante :

Thèse psychologique de Turing n° 2.1 :

> **La mise au jour d'un schème mécanique de pensée repose sur la possibilité d'une identification psychologique à une machine universelle.**

Il s'agit bien là d'une thèse psychologique dans la mesure où elle renverse le rapport qui existe entre l'intuition et ce qui n'est que son expression, à savoir la notion de machine, pour faire de la notion de machine une réalité indépendante dans laquelle l'intuition peut ou non s'investir.

3. Le rapport entre intuition et oracle :

Les champs d'application de l'intuition et du schème mécanique sont identiques, mais l'intuition produit des expressions mécaniques dans le virtuel qui ne sont pas toujours actualisables. Cette expression mécanique peut soit être schématisée (actualisée) et elle devient alors une machine-*a*, soit ne pas l'être et se constitue en machine-*o*. Dans ce dernier cas, l'intuition apparaît comme un oracle, c'est-à-dire comme une faculté psychologique exprimant le simple souhait d'une mécanisation possible sans que ce souhait puisse s'investir dans une machine réelle. On peut représenter cette conséquence de la thèse de Turing sous la forme du schéma suivant :

Thèse psychologique de Turing n° 2. 2 :

> L'existence de nombres inaccessibles par machine transforme

l'intuition en oracle, c'est-à-dire oblige le schème mécanique à faire retour sur sa source psychologique conçue comme simple souhait d'une mécanisation possible.

4. Le rapport de l'intuition-oracle au transfini :

Remarquons tout d'abord qu'*il y a* une théorie et un usage du transfini dans le cadre du modèle de la machine à état discret, contrairement à ce que l'on aurait pu le penser au départ. C'est ce qu'a montré à la fois l'analyse de la calculabilité des réels ainsi que celle de la représentation ordinale de la hiérarchisation des logiques.

Cette théorie du transfini apparaît clairement dans la différence que Turing opère entre ce qu'il appelle machine-*a* et machine-*o*, comme le montre le schéma du paragraphe 31. Ce schéma possède un caractère dynamique et, en tant qu'il institue un certain rapport au temps, il se laisse interpréter en termes psychologiques : dans le cas où le schème mécanique se heurte à une impossibilité (par exemple celle de la mise en liste de toutes les machines non-circulaires) et qu'il ne constitue qu'une machine-*o*, cette machine-*o* fait retour vers l'oracle en tant que faculté produisant des expressions mécaniques à jamais virtuelles.

Au vu de ce schéma, on peut se demander s'il n'y a pas ici une parenté avec la façon dont Brouwer et sa postérité intuitioniste constituaient le continu au moyen de la notion de suites de choix libres[150]. En particulier, la constitution d'une hiérarchie des logiques par le biais de la notion de machine-*o* qui parvient à effectuer une décision ou qui doit y renoncer en perdant la trace du schème de pensée mécanique ressemble à la notion de stérilisation chez Brouwer qui consiste à arrêter une suite quand on rencontre un “obstacle”, c'est-à-dire une contradiction qui empêche de poursuivre l'engendrement de la suite en

---

[150] Cf. I, 1, § 32. Cf. **J. Largeault**, *Intuition et intuitionisme*, op. cit., pp. 114 : «[…] le finitisme rendrait les mathématiques soit impossibles soit triviales. Cependant elles sont l'expérience d'une activité de conscience forcément finie. Cela étant, la seule conception tenable de l'infini doit reposer sur l'admission de processsus (de suites) indéterminés».

question[151]. C'est pourquoi l'expression employée par Brouwer pour caractériser le continu comme un "milieu en libre devenir" semble pouvoir s'appliquer à la façon dont le transfini apparaît au sein de la théorie générale du schème mécanique tel qu'elle apparaît dans le cadre de la thèse de Turing et ses implications psychologiques.

Une dernière remarque doit être faite concernant la thèse de Turing et son rapport à la thèse de Church. Alors qu'il est relativement aisé de caractériser la thèse de Church comme appartenant au domaine mathématique, il est moins facile de caractériser l'objet sur lequel porte ce que l'on a appelé "les thèses de Turing".

333. Interprétations des thèses de Turing : mathématique, physique, psychologie

Il est toujours possible d'interpréter la thèse de Turing dans un cadre mathématique comme un équivalent de la thèse de Church : Turing lui-même a montré dans un appendice à "On Computable Numbers …" que cette équivalence était mathématiquement démontrable[152]. Mais on a vu que la thèse de Turing avait des implications psychologiques que n'avait pas la thèse de Church et que son énoncé tendait même à occulter[153]. Quand on s'interroge sur la nature de ces implications psychologiques, une interprétation retient particulièrement l'attention : celle du physicien R. Penrose.

Penrose fait remarquer d'une part que c'est bien la maîtrise des nombres réels qui fait le fond du problème de la calculabilité[154] et d'autre part que les nombres réels semblent indispensables[155] à l'élaboration de toute physique

---

[151] Cf. **J. Largeault**, *Intuition et intuitionisme*, op. cit., pp. 113-115.

[152] Cf. **A. M. Turing**, (1936), "On Computable Numbers (…)" Appendix "Computability and effective calculability", dans [**M. Davis ed.**, *The Undecidable*, op. cit.], pp. 149-151.

[153] Cf. Première partie, chapitre 2, § 33.

[154] Cf. **R. Penrose**, *The Emperor's New Mind*, Oxford University Press, Oxford, 1989, p. 66.

[155] C'est au moins vrai jusqu'à présent. J. P. Delahaye fait remarquer que l'on essaye aujourd'hui de penser directement la physique en termes discrets : «(…) les soucis d'effectivité théorique (les nôtres) et ceux d'effectivité pratique convergent vers ce qui est une des voies les plus prometteuses pour la physique et qui permettra peut-être à terme de résoudre les difficultés épistémologiques graves qu'introduit l'usage du continu en physique.» **J. P. Delahaye**, "Formalisations

mathématique (qu'elle soit classique ou relativiste) parce qu'ils permettent de définir une géométrie, elle-même au fondement du cadre général spatio-temporel dans lequel des événements physiques peuvent recevoir une détermination. Le continu de l'espace-temps exige donc précisément que soit pris en compte ce type de nombres qui fait l'objet de l'enquête de Turing dans "On Computable Numbers…".

A partir de cette constatation, la "thèse de Penrose" consiste à dire que l'accès psychologique au non-calculable est *l'indice de la nature matérielle de l'esprit*. Ainsi le point de vue de Penrose lui permet-il de mener une enquête en termes *objectifs* sur le terme *subjectif* de la thèse de Turing : l'intuition. Alors qu'en se limitant au domaine du calculable, le fonctionnement cognitif de l'intuition devenait par le fait même très mystérieux, l'interprétation de Penrose permet, en accordant une réalité physique au non-calculable, d'analyser le fonctionnement cognitif de l'intuition en termes objectifs. Il lui paraît alors possible de rendre compte du fonctionnement cognitif de l'intuition en essayant de comprendre comment s'articule, en elle, ce qui relève et ce qui ne relève pas de la sphère du calculable. Penrose situe l'articulation en question à un point de passage entre le niveau physique macroscopique et le niveau physique microscopique[156]; ce point de passage physique lui paraît correspondre de plus à l'articulation dans la nature du niveau de description physique et du niveau de description biologique[157].

L'interprétation que Penrose accorde à la thèse de Turing est recevable :

---

mathématiques de la question : Le monde est-il récursif ?", op. cit., p. 201. Ces recherches sont encore trop programmatiques pour permettre de modifier l'interprétation classique de la physique en termes continuistes. La question au fond du débat porte sur la nature de la géométrie dont il n'est pas impossible qu'elle puisse se concevoir au sein d'un univers discret. Cf. pour ce débat, **J.-M. Salanskis**, *L'herméneutique formelle*, CNRS, Paris, 1991, p. 187-191.

[156] Cf. **R. Penrose**, "On Physics and Mathematics of Thought" in [*The Universal Turing machine*, **R. Herken ed.**, Oxford University Press, 1988], p. 519, dans lequel il prend l'exemple de la rétine du crapaud capable de réagir à un niveau macroscopique à l'action d'un seul photon.

[157] Il remarque par exemple que des organismes unicellulaires comme les paramécies sont dépourvus de système nerveux et sont cependant capables d'assurer leur contrôle moteur. Si l'on identifie le système nerveux à un système de traitement computationnel des informations en provenance de l'extérieur, alors, le contrôle moteur des paracémies ne peut pas être opéré par le biais d'un calcul. Cf R. Penrose, "Computability and Human Understanding", conférence faite au Department of Continuing Education, Oxford, le 6 juin 1993.

dans la mesure où il est question dans "On Computable Numbers…" de la calculabilité des réels et que cet ensemble numérique est bien au fondement de la géométrie nécessaire à l'élaboration d'une physique, il est légitime d'envisager, de la façon dont le fait Penrose, le projet d'une "intelligence artificielle".

Néanmoins, il nous a semblé qu'il y avait autre chose dans les thèses de Turing qu'un énoncé mathématique et physique parce qu'il nous est apparu que, dès le niveau de description mathématique et logique de la thèse de Turing, il était nécessaire de faire intervenir un niveau de description proprement psychologique. Comment dès lors interpréter le fonctionnement cognitif de l'intuition, ou, pour reprendre les termes de Turing, de "l'intuition-oracle", tout en accordant à la sphère psychologique une autonomie ? C'est en effet en récusant au psychologique un niveau de description autonome que Penrose parvient à analyser en termes objectifs le fonctionnement cognitif de l'intuition. Si, au contraire, on accorde au psychologique un niveau de description autonome par rapport à celui de la physique et de la biologie, comment réussir à penser leur articulation ?

Ce point de vue consiste à tenter de montrer que le continu, outre sa signification mathématique, géométrique et physique, paraît devoir revêtir également une signification psychologique spécifique. C'est ce que notre dernier chapitre a tenté d'établir au sein des écrits mathématiques et logiques de Turing et c'est cette signification psychologique que nous allons tenter de mettre au jour dans la deuxième partie en nous reposant sur d'autres textes de Turing. Le concept de continu apparaît alors comme un "fil d'Ariane" susceptible de mettre en rapport l'aspect physique et mental de la notion d'esprit et de circonscrire, par ce biais, l'objectivité propre au projet de l'intelligence artificielle. C'est ce que nous allons essayer de voir dans la seconde partie, intitulée "La logique dans la psychologie" et qui va consister à comprendre dans quelle mesure une analyse en termes logico-mathématiques peut rendre compte du niveau de description psychologique.

Deuxième partie

La logique dans la psychologie

«On pourrait, en interviewant les scientifiques, recueillir de nombreuses données sur la fascination qui sous-tend leur travail de recherche. L'interprétation de ces données serait délicate mais permettrait peut-être une meilleure compréhension psychologique du processus de découverte scientifique. Les savants atteints de folie ou de sénilité seraient particulièrement intéressants à étudier, à cause de la plus grande transparence de leurs motivations».

David Ruelle, *Hasard et chaos*, p. 277, note 2.

Introduction

L'investigation psychologique que nous allons mener dans cette deuxième partie exige un autre type d'intelligibilité que l'investigation formelle menée dans la première. Pourquoi ? Il faut, pour le comprendre, situer les acquis de la première partie dans la perspective d'une théorie de la psychologie.

1. Justification de l'usage d'une méthode informelle

11. L'idéalité des objets mathématiques et la thèse de Turing

En tant qu'énoncé mathématique, la thèse de Turing entre dans le cadre de ce que Poincaré appelait la connaissance mathématique qui repose, selon lui, sur une «contradiction insoluble» : d'une part, on ne peut pas réduire les théorèmes mathématiques à l'application réitérée des règles de la logique sur les axiomes sans faire de la mathématique elle-même «une immense tautologie», ce qui semble peu plausible; d'autre part, on ne peut pas non plus concevoir comment une vérité mathématique pourrait receler plus que ce qu'on y a mis, à savoir les principes de la déduction logique appliqués aux axiomes et eux seuls[158]. C'est cette contradiction qui fait que les objets mathématiques apparaissent toujours en excès par rapport à leur détermination, excès qui demande à être indéfiniment retravaillé pour être maîtrisé. A l'origine de cet excès se trouve le processus d'idéalisation qui constitue les objets mathématiques tels que la droite ou l'algorithme en objets abstraits. C'est l'aspect abstrait de ces entités qui exige que soit institué un processus d'éclaircissement indéfini. Par définition en effet, le plan abstrait est en rupture par rapport au plan empirique de la nature : aucun objet

[158] **H. Poincaré**, *La Science et l'hypothèse*, Flammarion, Paris, 1902, p. 31.

mathématique n'est réalisé dans la nature et c'est seulement sur le plan de l'idéalité qu'il a une existence. C'est ce statut d'existence apparemment si paradoxal que décrivait Poincaré quand il mettait au jour une contradiction fondatrice dans la connaissance mathématique.

La profonde originalité de la thèse de Turing, en tant qu'elle est un énoncé relevant de la discipline abstraite des mathématiques, vient de ce que *le processus de constitution du plan de l'idéalité n'est plus caché* puisque la thèse ne fait rien d'autre que décrire ce processus de constitution en effectuant le passage de l'informel au formel. C'est pourquoi la thèse de Turing possède ce statut paradoxal d'être un énoncé idéal portant sur l'acte psychologique d'idéalisation.

Aussi cet acte psychologique de projection sur le plan du formel apparaît-il comme un objet mathématique soumis à la "contradiction" dont parlait Poincaré[159], contradiction qui, dans ce contexte, s'exprime sous la forme suivante : le processus psychologique décrit idéalement (mathématiquement) par la thèse de Turing ne contient rien d'autre que du mécanisme et en même temps, il doit contenir plus que lui. Cet énoncé apparemment paradoxal est la manifestation de l'appartenance de la thèse de Turing au registre idéal de la connaissance mathématique. Mais cette idéalité est d'un genre très particulier puisqu'elle renvoie à l'acte d'idéalisation lui-même, à savoir le passage de l'informel au formel. C'est cet acte d'idéalisation qui rend possible de projet de l'intelligence artificielle.

12. La réflexion du plan de l'idéalité et le projet d'intelligence artificielle

Cette réflexion du plan de l'idéalité sur lui-même rend en effet possible la constitution d'une connaissance formelle de l'acte d'idéalisation et c'est ainsi que nous définirons le projet d'intelligence artificielle. Cette définition permet d'expliquer son projet scientifique concret, que D. Marr décrivait en ces termes

---

[159] On trouve dans l'article de R. J. Nelson des remarques semblables à celles de Poincaré à propos de la thèse de Church : *a priori* du point de vue méthodologique, elle paraît *a posteriori* du point de vue mathématique. L'auteur, tout en employant un vocabulaire kantien (il parle en particulier de «jugement analytique *a priori*») et bien que son analyse s'y prêterait, ne décrit pas la thèse de Church comme un jugement synthétique *a priori*. Cf. **R. J. Nelson**, "Church Thesis and Cognitive Science", Notre-Dame Journal of Symbolic Logic, XXVIII-4, 1987, p. 581-614.

[160]:

> **«L'intelligence artificielle est l'étude des problèmes complexes de traitement de l'information. Le but de cette discipline est d'identifier des problèmes de traitement de l'information intéressants et solubles et de les résoudre. […] De façon stricte, un résultat en intelligence artificielle consiste à isoler un problème de traitement de l'information particulier, à formuler une théorie computationnelle qui lui corresponde, à construire un algorithme qui l'implémente et à démontrer pratiquement que l'algorithme est efficace».**

On voit que la place de la notion d'algorithme est fondamentale dans cette définition : en fait, l'intelligence artificielle consiste à montrer que toute notion psychologique de nature informelle peut se ramener au cas décrit par la thèse de Turing et qu'elle peut donc faire l'objet d'une analyse en termes formels.

Résoudre un problème en intelligence artificielle consiste donc à identifier la présence d'un traitement calculatoire à différents niveaux de description, niveaux qui vont du logico-mathématique au physique. Cette attitude méthodologique se justifie donc facilement dans ce domaine quand elle permet de construire un modèle formel d'une activité psychologique jusqu'alors décrite en termes informels.

Cependant, il faut remarquer qu'elle utilise les concepts mathématiques d'une façon radicalement différente de celle qui était à l'œuvre dans la thèse de Turing parce qu'elle envisage les termes de celle-ci comme "à rebours": c'est en effet la notion formelle de machine qui vient s'appliquer aux notions informelles de la psychologie. Une conséquence immédiate en découle. Dans le cas de l'intelligence artificielle, le concept mathématique de machine a *d'emblée* un statut formel - puisqu'il sert de modèle - sans que soit repensé son origine psychologique, c'est-à-dire sans que soit retracé le processus d'idéalisation décrit dans la thèse de Turing; dans le cas de la théorie mathématique en revanche, la notion d'algorithme ne devient formelle que par l'intervention d'un acte psychologique qui en assure la traduction sous l'aspect de la notion de machine. Aussi, dans le cas de l'intelligence artificielle, le concept de machine peut-il servir

---

[160] **Marr D.**, "Artificial Intelligence : a personal view" dans [**D. Partridge** et **Y. Wilks eds.**, *The Foundations of Artificial Intelligence*; *a sourcebook*, Cambridge University Press, Cambridge, 1990], p. 97.

de forme pouvant s'appliquer à tout contenu, y compris à la source psychologique dont le concept émane. C'est pourquoi l'intelligence artificielle tend à ne pas prendre en considération le fait que les notions psychologiques apparaissent d'abord, à celui qui les exprime "de l'intérieur", de façon informelle. Il devient dès lors possible de considérer comme non-pertinent d'un point de vue scientifique toute enquête informelle sur l'acte psychologique qui effectue le passage de l'informel au formel. De ce point de vue, la démarche de l'intelligence artificielle ressemble à celle d'une modélisation en physique dans laquelle les concepts et modèles mathématiques sont envisagés comme des formes qui viennent s'appliquer à un contenu matériel qui leur est étranger : aussi tend-on à oublier le fait que ces formes mathématiques sont elles-mêmes des expressions provenant d'une source psychologique intuitive et qu'elles ont donc tout d'abord eu à jouer le rôle de contenu par rapport à ce contenant psychologique premier.

Ce dernier point semble, au premier abord, permettre une critique à l'égard du projet d'intelligence artificielle. En tant qu'elle apparaît comme la possibilité d'une mécanisation de l'acte psychologique qui va de la notion informelle d'algorithme à la notion formelle de machine, on peut se demander si cette démarche n'implique pas en elle-même de produire un résidu non-mécanisable inaccessible, résidu qui provient de ce qu'elle cherche à modéliser par la notion de machine un processus psychologique qui n'est pas intrinsèquement mécanique puisqu'il décrit le passage de l'informel au formel. Ce résidu inaccessible au mécanisme ne condamne-t-il pas d'emblée le projet d'une formalisation générale des processus psychologiques ?

En fait, cette critique n'est pas pertinente : l'intelligence artificielle a un statut d'idéalité et comme telle, elle ne vise pas la connaissance de l'acte psychologique d'idéalisation mais uniquement sa modélisation. La mécanisation de la source psychologique informelle est un processus indéfini qui ne se distingue pas de ce point de vue d'autres types d'enquête scientifique et en premier lieu des mathématiques. Si l'on revient en effet à la façon dont Poincaré décrivait l'idéalité des objets mathématiques, on comprend mieux la situation apparemment paradoxale de l'intelligence artificielle qui, en prenant les termes de la thèse de Turing “à rebours”, semble confondre dans un même plan d’intelligibilité l’expression mécanique et la source psychologique de cette expression. On a vu qu’adapté au contexte qui est le nôtre, le dilemme de Poincaré s’énonce ainsi : d’une part, le processus psychologique décrit idéalement (mathématiquement) par la thèse de Turing ne contient rien d’autre que du mécanisme et d’autre part, il doit contenir plus que lui. Cette contradiction ne demande pas à être résolue puisqu’elle reflète seulement l’existence d’un domaine idéal. Elle permet au contraire de justifier la place de l’intelligence artificielle.

Schématiquement, on peut dire que l’intelligence artificielle se situe sur la première branche du dilemme de Poincaré : en prenant les termes de la thèse de Turing “à rebours”, l’intelligence artificielle tente de rapporter à des principes purement formels l’acte psychologique d’idéalisation. Ce projet scientifique est viable et a déjà montré sa vitalité, en particulier dans la théorie de la vision[161]. Mais cette caractérisation de l’intelligence artificielle n’a de sens que si l’on prend en compte l’autre branche du dilemme, qui souligne au contraire que l’idéalité de l’acte psychologique d’idéalisation implique aussi qu’il ne puisse pas être rapporté à une source seulement logique à partir de laquelle il serait possible de le déduire. De ce dernier point de vue, il faut donc tenter d’expliquer ce processus d’idéalisation non pas du point de vue de l’idéalité, tâche de l’intelligence artificielle, mais du point de vue du processus lui-même, avant que ne soit constitué une sphère de l’idéalité à part entière. Or dans ce cas, on ne peut plus avoir recours au formel puisqu’on essaye de penser son émergence : *il est donc nécessaire de ce point de vue de faire une place pour une analyse informelle de l’acte psychologique d’idéalisation*.

C’est ce que nous allons tenter d’étudier dans cette deuxième partie en

---

161 La théorie de la vision telle qu’elle a été constituée par D. Marr et T. Poggio constitue une science à part entière que ces auteurs ont appelé “théorie de l’optique inverse”. Elle distingue deux étapes temporelles, la vision primaire et la vision de haut-niveau et trois niveaux d’intelligibilité, le niveau computationnel, le niveau algorithmique et le niveau physique. La difficulté spécifique provient de l’imbrication de ces trois niveaux qui vont du mathématique au physique et qui requièrent une analyse en termes de représentations et de contraintes physiques. Cf. **A. Hulbert** et **T. Poggio**, “Making machines (and artificial intelligence) see” dans [**S. T. Graubard ed.**, *The Artificial Intelligence Debate, False Starts, Real Foundations*, Cambridge, Mass : MIT Press, 1988], p. 237.

usant d'un autre type d'intelligibilité que celui offert par l'investigation formelle puisque celle-ci, dans la mesure où elle est formelle, ne permet pas, à elle seule, de servir le but de notre enquête. Il ne s'agit donc pas de renoncer à l'investigation formelle mais seulement de tenter d'en situer la place dans une perspective plus générale concernant le statut d'idéalité de la thèse de Turing. Plus concrètement, il ne s'agit donc pas d'étudier comment l'informel et le formel s'articulent - c'est la thèse de Turing qui rend raison de cette articulation - puisque dans ce cas, on présuppose l'existence des termes à mettre en rapport mais de rendre compte de l'activité psychologique de production d'un domaine du formel, celui du mécanique.

L'objet de cette enquête est donc la description de l'acte psychologique qui va de l'informel au formel et qui constitue l'idée de mécanisation. C'est ce "souhait de mécanisation" dont il faudrait essayer de rendre compte. G. Kreisel le décrit en ces termes [162]:

> **«Qu'y a-t-il de si merveilleux dans la formalisation ? On s'est battu à plate couture pour trouver une réponse. Une seule sera prise en considération ici. C'est le présupposé tacite d'un besoin éthéré - ici satisfait par le biais de la formalisation - pour une norme ultime de précision; un présupposé tacite répandu non seulement dans la recherche sur les fondements des mathématiques mais partout ailleurs dans la culture occidentale».**

Kreisel ne tente pas de rendre raison de façon plus précise de ce désir éthéré qui traverserait la culture occidentale. Cette remarque reste donc purement métaphysique : en accordant une influence causale à une entité aussi obscure que la "culture occidentale", on ne voit pas quel sens accorder, sinon ironique, à ce "désir éthéré". Or c'est précisément ce "désir éthéré" qu'il faut essayer de décrire en termes non métaphysiques.

2. Méthode employée

La méthode que nous emploierons dans la suite consiste non pas à prendre les termes de la thèse de Turing "à rebours" comme le fait l'intelligence artificielle, mais à tâcher de comprendre comment le formel peut se constituer à

---

[162] **G. Kreisel**, "Review of K. Gödel 'Collected Works, I''", Notre Dame Journal of Formal Logic, volume 29, n°1, Winter 1988, p. 165.

partir de l'informel quand on ne présuppose pas l'existence du premier. Il faut, pour ce faire, établir une distinction entre deux modes différents dans l'appréhension des symboles.

21. Symboles cognitifs et symboles praxiques dans la thèse de Turing

Comme le faire remarquer D. Widlöcher, on doit distinguer deux types de symboles quand on s'interroge sur la nature des facultés psychologiques au fondement de la constitution des représentations [163]:

> **«Une distinction qui semble très importante pour notre propos, tient à la fonction du symbole et non à sa forme. C'est elle qui oppose le symbole destiné à représenter un objet ou un état du monde, comme un symbole chimique ou linguistique, au symbole destiné à donner sens à un acte, comme le signe de la croix, le fétiche sexuel et, en général, les symboles des rêves et des jeux. Il semble utile de marquer une différence radicale entre l'activité symbolique destinée à la construction d'un système de représentations et l'activité destinée à exprimer un acte. Je propose que nous parlions de symboles cognitifs pour définir ceux qui ont ainsi pour fonction de représenter des éléments d'information et de permettre leur traitement. […] Je propose que nous dénommions symboles praxiques ces objets ou ces signes qui ont pour fonction de figurer un acte».**

C'est évidemment les rapports qu'entretiennent ces deux types de symboles qu'il faut réussir à déterminer puisque la psychologie ne se limite pas à la sphère du cognitif telle qu'elle a été définie par D. Widlöcher. Pour lui, le rapport en question est d'opposition radicale. Les symboles cognitifs seraient ceux que l'on manipule dans les connaissances scientifiques tandis que les symboles praxiques se rapporteraient au fonctionnement inconscient tel qu'il apparaît dans les lapsus ou dans les rêves[164]. Cependant, si l'on se rapporte à la thèse de Turing et au fait qu'elle fait usage en elle-même de l'acte psychologique qu'elle décrit, il semble qu'on ne puisse pas distinguer aussi radicalement un fonctionnement symbolique purement cognitif d'un fonctionnement symbolique de type praxique, c'est-à-dire qui viserait une "théâtralisation". En supposant une distinction

---

[163] **D. Widlöcher**, "Croire en l'inconscient", Nouvelle revue de psychanalyse, n°48, automne 93, Gallimard, Paris, 1993, pp. 93-113.

[164] Cf. **D. Widlöcher**, "Croire en l'inconscient", Nouvelle revue de psychanalyse, n°48, op. cit., p. 105 : «Quand on considère l'activité symbolique qui intervient dans la construction des connaissances scientifiques, il est clair qu'elle est constituée par des symboles cognitifs qui s'inscrivent dans une structure de code».

radicale entre les deux modes de fonctionnement, c'est l'acte psychologique à la source de la capacité de calcul qui, par le fait même, devient obscur puisque cet acte, par son existence même, remet en question la distinction entre cognitif et praxique. La thèse de Turing et ses conséquences psychologiques ont précisément montré que la distinction de deux types de fonctionnement symbolique ne pouvait être marquée de façon radicale. Mais la thèse de Turing n'a pas résolu la question de savoir quels étaient les rapports existant entre les deux modes de fonctionnement symbolique. La question qu'il faut étudier est donc celle de l'articulation des deux fonctionnements du point de vue du processus psychologique d'idéalisation.

22. Le jeu comme activité symbolique

Il y a une activité humaine qui se prête à l'étude de cette articulation : c'est l'activité de jeu. Le jeu possède en effet un aspect cognitif et un aspect praxique.

Son aspect cognitif vient de ce que tout jeu obéit à des règles que l'on peut qualifier de formelles dans la mesure où elles ne répondent à aucune signification qui leur préexisterait. Par exemple, il n'y a pas de raison, aux échecs, pour que le fou se déplace de biais : le fait qu'il se déplace ainsi relève d'une convention.

Son aspect praxique vient de ce qu'un jeu, quand il est en train d'être joué, ne vise que son propre accomplissement. Le jeu semble ainsi dénué de tout but qui lui serait extérieur puisqu'il ne vise la réalisation d'aucune action réelle; cela ne veut pas dire qu'il n'y ait pas d'actions dans les jeux (à preuve les jeux sportifs) mais les actions qui y sont menées ne visent pas la réalisation d'un but mais seulement la représentation de cette réalisation. Ainsi aux échecs, faire échec et mat ne fait que représenter la victoire sur l'adversaire et ne vise pas une victoire réelle[165]. Le but d'un jeu n'est donc pas de parvenir à agir mais seulement de se représenter une action.

De ce point de vue, le jeu entretient avec une autre notion qui n'a de sens

---

[165] On peut évidemment imaginer un jeu où la victoire imaginaire serait aussi suivie d'une victoire réelle. Aux échecs par exemple, la victoire par échec et mat pourrait consister en la mise à mort de l'adversaire, si l'on respecte l'étymologie arabe de “mat” [mort]. A l'évidence, cette seconde victoire ne fait pas partie du jeu dans la mesure où elle résulte d'une convention qui n'est pas liée aux règles de celui-ci.

que par rapport au domaine de la représentation un rapport certain : c'est la notion d'apprentissage. L'apprentissage ne vise en effet aucune réalisation réelle : même si l'on souhaite apprendre *en vue* de réaliser telle ou telle action et que l'apprentissage consiste le plus souvent à effectuer l'action que l'on veut réaliser, l'apprentissage en tant que tel consiste à *mimer* cette action en vue de la réaliser ultérieurement au mieux. L'apprentissage se situe donc avant toute réalisation et la réalisation qu'elle effectue se situe de ce fait dans le domaine de la représentation, comme c'est le cas du jeu.

221. La notion de représentation dans le jeu et l'apprentissage

Comme l'a remarqué Norbert Wiener[166], tout système organisé peut être considéré comme transformant un message d'entrée en message de sortie suivant un principe de transformation. Si le principe de transformation est soumis à un critère permettant de mesurer la valeur de la performance du système et si le système organisé en question est réglé en vue d'améliorer ses performances par rapport à ce critère, on dit que le système *apprend*. Or il est possible de se représenter ce type de système organisé par le moyen de la notion de jeu. Wiener fait ainsi remarquer [167]:

> **«Un type très simple de système possédant un critère de performance facile à interpréter selon des règles fixes est un jeu dans lequel le critère de performance est la victoire telle qu'elle a été définie par ces règles».**

L'apprentissage pour les joueurs consiste donc à améliorer leurs chances de victoire. Quelle est la nature de cet apprentissage ? Il peut être considéré comme l'adoption d'une stratégie optimale. A l'évidence, la stratégie la meilleure consisterait à savoir gagner pour toutes les parties. Ce type de stratégie a été défini axiomatiquement par von Neumann[168] et, en droit, *celle-ci est valable pour tous les jeux*. Elle consiste à suivre une fois pour toutes et quelle que soit la stratégie de l'adversaire (parce que l'on sait répondre à tous ses coups possibles) le plan

---

[166] Cf. **N. Wiener**, *God and Golem, Inc.*, MIT Press, Cambridge, Massachusetts, 1964, p. 14.
[167] **N. Wiener**, *God and Golem, Inc.*, op. cit., p. 14.
[168] Cf. **B. Saint-Sernin**, *Les Mathématiques de la Décision*, PUF, Paris, 1973, p. 175.

d’action qui a été adopté au départ, c’est-à-dire avant même le début du jeu. On appellera cette stratégie la “stratégie d’omniscience”.

En fait, pour des raisons pratiques tenant à l’explosion combinatoire qu’implique la mise à l’étude de toutes les réponses possibles de l’adversaire à un coup donné, peu de jeux peuvent être joués au moyen d’une stratégie d’omniscience. Aussi, dès que l’on a affaire à un jeu un peu “compliqué” comme les échecs, cette stratégie n’est-elle plus praticable, au moins tant que les capacités de stockage des coups possibles sont relativement limitées. Dans ce cas, il faut adopter une tout autre attitude qui prend en compte la stratégie de l’adversaire et permet de réagir par rapport à celle-ci. Borel faisait remarquer à ce propos [169]:

> **«[…] La connaissance de la psychologie de l’adversaire doit, à *chaque instant,* entrer en ligne de compte pour *modifier* les règles de conduite qu’on adopte».**

Cette stratégie, que nous appellerons “pragmatique” par opposition à la “stratégie d’omniscience”, repose sur la reconnaissance de l’existence d’un adversaire et de son intériorité : elle a donc pour objet la *psychologie* de l’adversaire et la victoire passe par l’analyse mathématique de cet objet dont on postule l’existence, la psychologie *d’autrui*.

On peut donc dire que les deux stratégies se distinguent par la réalité qu’elles attribuent à la psychologie.

La première, en éliminant la question de la réalité de la psychologie, supprime la validité de toute enquête psychologique qui n’a dès lors plus d’objet propre puisque la “stratégie d’omniscience” ne prend en compte la stratégie de l’adversaire que par l’introduction d’un changement aléatoire dans sa propre stratégie. La seconde se fonde au contraire sur la reconnaissance de la réalité de la psychologie de l’adversaire et de la validité d’une enquête mathématique la concernant. Bref, dans le premier cas, on peut *faire comme si* l’adversaire n’avait pas de psychologie, tandis que dans le second, on accorde une *réalité* à un objet psychologique dont on ne peut que postuler l’existence.

Ainsi la notion de jeu permet-elle de prendre conscience du fait qu’il y a

---

[169] **E. Borel**, Note du 19 déc. 1921, *Compte-Rendus de l’Académie des Sciences*, t. 173, p. 1304-1305, cité dans [**B.Saint-Sernin**, *Les Mathématiques de la Décision*, op. cit.], p. 176.

un rapport entre le débat sur la nature de l'apprentissage dans les systèmes organisés telle qu'elle apparaît dans le modèle du jeu et le débat sur la nature de la psychologie. Tâchons d'expliciter ce rapport par un exemple.

Pour un jeu "compliqué" comme les échecs, la question que l'on peut poser est celle-ci : dans le cas où l'un des adversaires est une machine universelle de Turing dont les capacités de stockage sont par définition infinies et qui est correctement programmée compte tenu des règles du jeu, quel type de stratégie doit être employé pour parvenir à ce que la machine gagne ? Comme on vient de le voir, c'est le problème du statut de la *psychologie* qui est le réel enjeu de la question. Aussi peut-elle se traduire sous cette forme : quelle place y a-t-il pour la psychologie dans un jeu où l'un des adversaires est omniscient ? Ou encore : quelle psychologie attribuer à un système organisé du type de celui d'une machine universelle de Turing ?

Avant d'aborder la réponse à cette question qui va nous occuper tout au long de cette partie, remarquons que son enjeu est aussi un héritage de la tradition philosophique. Comme l'a noté Norbert Wiener, passant outre à toute timidité positiviste, le problème de l'apprentissage de machines qui apprennent à jouer à des jeux, est de nature *théologique* : c'est le problème du rapport entre le créateur et la créature qui s'y trouve en fait posé. Wiener remarquait en effet à propos des machines et de ceux qui les inventait[170]:

> **«En construisant des machines avec lesquelles il joue, l'inventeur s'est arrogé la fonction d'un créateur limité, quelle que soit la nature de l'appareil à jouer qu'il a inventé. Ceci est particulièrement vrai dans le cas de machines pouvant jouer et qui apprennent par expérience».**

En jouant, le créateur du jeu ne fait qu'obéir aux règles qu'il a lui-même inventées. De ce point de vue, il joue en fait avec sa propre création et découvre donc, par le biais d'une manipulation de symboles praxiques, ce qu'il ne sait pas à propos de lui-même. Le jeu qu'il a inventé apparaît donc comme une expérience indirecte de sa propre faculté psychologique d'invention. Aussi le modèle du jeu permet-il de mener, d'un point de vue général, une enquête sur ce que l'on entend

---

170 **N. Wiener**, *God and Golem, Inc.*, op. cit., p. 17.

par invention, parce que c'est par le biais du jeu que l'on peut parvenir à s'en faire une idée. Le jeu entretient ainsi avec l'explicitation de la nature de l'esprit un rapport profond.

Un texte de la tradition philosophique décrit précisément le rapport du créateur à la créature par le modèle du jeu : c'est le pari de Pascal [171]. Dans ce texte, le rapport de la créature au créateur prend la forme d'une enquête de nature probabiliste[172]. En se plaçant du point de vue de la créature, le type de stratégie adoptée par Pascal est une stratégie de type "pragmatique" qui tend à fonder son plan d'action sur la réalité de la psychologie de l'adversaire, en l'occurrence le créateur omniscient. Ce faisant, Pascal confère à "l'adversaire" une psychologie réelle. Mais selon ce que nous disions de la nature des jeux, c'est par ce biais qu'il se donne aussi les moyens d'étudier le fonctionnement de *sa* psychologie.

Nous retrouverons ce type d'argument dans le cadre de la réflexion menée par Turing sur la "psychologie" qu'il faut prêter à la machine universelle quand on l'étudie dans une situation d'interaction entre joueurs propre à un jeu spécifique, et plus généralement, sur le type de théorie psychologique que l'on peut construire à partir du projet de l'intelligence artificielle.

Ainsi la notion d'apprentissage pour un système organisé et celle de psychologie pour l'être humain peuvent-elles s'expliciter l'une l'autre, si l'on parvient à construire un modèle de jeu dont les résultats pourront être interprétés dans la perspective d'une enquête sur la nature de l'invention.

222. Utilisation du modèle du jeu par Turing

Turing ne s'y était pas trompé lui qui a inventé un jeu tout à fait particulier, le "jeu de l'imitation", pour étudier l'articulation du plan informel et du plan formel, selon le schéma de la thèse de Turing.

On peut assimiler , en suivant la définition proposée par N. Wiener, le "jeu de l'imitation" à un système organisé susceptible d'apprentissage : un message

---

[171] Cf. **Pascal**, *Pensées*, "Infini-Rien", n° 418 (éd. Lafuma) ou 233 (éd. Brunschvicg).

[172] Ian Hacking a montré qu'il fallait voir dans l'argument du pari l'acte de naissance du calcul des probabilités. Cf. **I. Hacking**, *The Emergence of Probability*, Cambridge University Press, Cambridge, 1975, chapitre 8, pp. 63-72.

d’entrée de nature informelle - les questions de l’interrogateur - est transformé en un message de sortie de nature informelle par un principe de transformation constitué d’une boîte noire dont on ne sait pas s’il s’agit d’un être humain (objet informel) ou d’un ordinateur “digital”[173] (objet formel). Dans ce cas, le montage expérimental réalisé dans le jeu permet d’étudier l’articulation du plan informel et du plan formel, puisque le but du jeu consiste à essayer de deviner la nature de la boîte noire en question.

C’est donc ce jeu qu’il faut tenter d’analyser pour légitimer notre point de vue sur la place à accorder à un “modèle computationnel de l’esprit”[174]. Le jeu de l’imitation est décrit par Turing dans un article qui date de 1950, “Computing Machinery and Intelligence”[175]. Cet article a servi et sert encore de “charte” pour l’intelligence artificielle. Placé en tête de nombreuses anthologies[176], il a été abondamment commenté depuis sa parution et on le considère habituellement comme le premier article d’intelligence artificielle. Son analyse ne nous paraît pourtant pas épuisée : il recèle de multiples pièges, obscurités, ironies, mots à double sens et demi-confidences autobiographiques qu’il paraît impensable de justifier par une analyse en termes purement formels. Ce qui nous semble être décrit dans cet article est l’*invention* du modèle computationnel de l’esprit et non pas du tout sa *viabilité*, contrairement au but déclaré de l’article. C’est pourquoi le jeu de l’imitation nous paraît décrire d’une part l’itinéraire *psychologique* qui va de l’informel au formel pour le cas particulier de l’individu Turing - et non pas la

---

173 Nous nous sommes habitués en français à cet anglicisme subreptice que représente le terme de “digital”. Il n’y a pas d’équivalent français direct au terme anglais [*digit*], puisque “chiffrique” n’est pas reçu. “Digital” étant de racine latine, il ne semble pas choquant de l’employer en français, d’autant plus que l’usage courant semble l’avoir assimilé.

174 L’expression de “modèle informationnel de l’esprit” serait sans doute plus heureuse en français mais les termes de “computationnalisme” et de “computationnel” formé en anglais à partir du terme de [*computer*], semblent faire aujourd’hui l’unanimité parmi la communauté cognitive. La racine de l’expression étant latine, elle semble pouvoir faire retour en français sans dommage.

175 Cf. **A. M. Turing**, “Computing Machinery and Intelligence”, Mind, Vol. LIX. n° 236, pp. 433-460.

176 On peut citer : **A. R. Anderson ed.**, *Minds and Machines*, Prentice-Hall, Englewood Cliffs, N. J., 1964; traduction française sous le titre *Pensée et Machine* Champ Vallon, Seyssel, 1983, pp. 39-67; **Butterworth ed.**, *Key Papers : Brain Physiology and Psychology*, University Park Press, Manchester, England, 1967; **D. Hofstadter** et **D. C. Dennett eds.**, *The Minds’I*, Basic Books, New York, 1981; traduction française sous le titre*Vues de l’esprit*, InterEditions, Paris, 1987, pp. 61-104 et **M. Boden ed.,** *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990, pp. 40-66.

formalisation du domaine psychologique en général - et d'autre part la psychologie que l'on peut attribuer au modèle de la machine de Turing dans le cadre d'une situation ludique d'interaction entre joueurs.

Nous allons donc tenter de montrer comment le modèle du jeu tel qu'il a été élaboré par Turing peut permettre de fonder un autre type de réflexion que celle de l'intelligence artificielle sur l'activité psychologique du calcul et de situer ainsi l'intelligence artificielle à la place qui lui revient dans l'enquête sur la nature et le fonctionnement de l'esprit. Il s'agit en quelque sorte de prendre l'attitude méthodologique de l'intelligence artificielle "à rebours" en retrouvant ainsi le sens originel de la thèse de Turing : au lieu de modéliser mathématiquement une activité psycho-physique, à la manière décrite par D. Marr, il faut essayer de comprendre comment le psycho-physique peut donner lieu à une idéalisation sur le point précis de la notion d'algorithme. C'est pourquoi l'articulation du formel et de l'informel recoupe en fait la distinction entre l'idéalité des objets mathématiques et la réalité du monde physique. On verra que c'est dans ce débat que réapparaît la question de la nature du continu parce que c'est par ce biais que Turing caractérise le monde matériel.

Chapitre I

Le jeu de l'imitation

Turing, dans "Computing Machinery and Intelligence", se propose de réfléchir à la notion d'intelligence en élaborant un jeu, le "jeu de l'imitation", permettant de juger de la présence ou de l'absence d'un comportement intelligent chez les êtres humains et les ordinateurs digitaux.

Pourquoi mettre l'accent sur les ordinateurs digitaux et non pas sur d'autres types de machines ? On peut invoquer deux raisons, la première théorique et la seconde historique.

Premièrement, parce que les ordinateurs digitaux sont les paradigmes de toute machine. En tant qu'ils sont les matérialisations au sein du monde physique des machines de Turing définies dans "On Computable Numbers …" et plus particulièrement de la notion de machine universelle, ils peuvent imiter n'importe quelle opération et donc exécuter les instructions de n'importe quelle machine de Turing, au moins du point de vue logique du contrôle des opérations[177]. Ces matérialisations de machines de Turing n'en sont évidemment que des modèles approchés puisque celles-ci disposent d'un ruban infini sur lequel opérer des calculs, alors que toute machine au sein de l'univers physique ne possède qu'une

[177] L'ordinateur est bien le paradigme logique de toute machine mais ne vise pas à remplacer les machines exécutant des tâches impliquant un mouvement physique : un ordinateur ne produit pas du mouvement comme une machine à vapeur, même s'il peut simuler les échanges thermodynamiques indispensables au fonctionnement d'une machine de ce type.

capacité de mémoire limitée (mais qu'il est toujours possible d'augmenter si l'on en possède les moyens techniques).

Deuxièmement, une autre raison, plus contingente, vient s'ajouter à la première et qui explique pourquoi la réflexion de Turing se situe désormais non seulement dans le cadre théorique des machines "de papier"[178] telle qu'il les avait imaginées en 1936 mais aussi dans le cadre concret des ordinateurs réels : un peu plus de deux ans avant la parution de "Computing Machinery and Intelligence" qui date d'octobre 1950, le premier ordinateur digital au monde a commencé à fonctionner à l'université de Manchester, précisément le 21 juin 1948[179]. C'est donc dans la continuation de cette première mondiale à laquelle Turing a participé à la fois d'un point de vue théorique depuis son article de 1936 et d'un point de vue pratique au *National Physical Laboratory* à partir de 1945 puis à l'université de Manchester elle-même à partir d'octobre 1948, qu'il rédige son article de nature philosophique.

Le jeu de l'imitation apparaît comme une expérience de pensée qui vise à convaincre le lecteur de l'article que, du point de vue de ce que l'on a coutume d'appeler un "comportement intelligent", il n'est pas possible de prêter ce type de comportement à l'être humain dans le jeu sans également le prêter aux ordinateurs digitaux convenablement programmés pour le jeu. La conclusion que Turing tire de l'expérience du jeu revêt ainsi la forme d'une implication logique de type "si

---

[178] Par "machine de papier", Turing entend l'effectuation des instructions par un opérateur humain. Cf. **A. M. Turing**, "Intelligent Machinery", op. cit., p. 34 : «Il est possible de produire l'effet d'une machine à calculer en écrivant l'ensemble des règles de procédure et en demandant à un homme de les effectuer. La combinaison d'un homme et d'instructions écrites sera appelée "Machine de papier"».

[179] Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, Unwin Paperbacks, London, 1983, p. 385 et p. 392. Il s'agissait d'une machine à architecture von Neumann (c'est-à-dire composée de cinq éléments fondamentaux : unité arithmétique; unité de contrôle; mémoire; unité d'entrée et unité de sortie) possédant une mémoire de 1K sous la forme d'un tube à vide. Le programme avait été rédigé par Tom Kilburn et consistait à trouver le plus grand facteur d'un nombre entier. Turing n'a participé ni à la construction effective de ce premier ordinateur appelé par les ingénieurs de Manchester "MADM" [*Manchester Automatic Digital Machine*] et plus communément la "machine-bébé" [*baby-machine*] à cause de la taille de sa mémoire, ni à la rédaction des premiers programmes. Nommé à Manchester en mai 1948, il n'a rejoint son poste qu'à la rentrée universitaire suivante, en octobre 1948. Il avait cependant envoyé courant juin son premier programme (effectuation d'une division) aux responsables du département. Il se trouve que ce programme comportait une erreur. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 560 note 6.

… alors”.

Exposons maintenant dans le détail le but et les règles du jeu de l’imitation.

1. But du jeu de l’imitation

Le but visé par Turing dans cet article est de montrer que, dans le cadre spécifié par le jeu de l’imitation, la notion d’intelligence peut être attribuée aussi bien aux êtres humains qu’aux ordinateurs digitaux construits sur le plan de la machine de Turing universelle. Deux problèmes majeurs se posent; le premier concerne la notion même d’intelligence et le second, la différence physique qui existe entre les êtres humains et les ordinateurs digitaux.

11. La notion d’intelligence

Pour pouvoir attribuer celle-ci aux êtres humains et aux ordinateurs digitaux, il faudrait posséder une définition de l’intelligence, ou du moins un trait permettant de la caractériser. Comment réussir à caractériser la notion d’intelligence ? Négativement, Turing refuse toute définition ou description qui ferait appel à une intériorité impalpable parce que celle-ci est par le fait même inaccessible. Il coupe donc court à la tradition qui consiste à faire appel à une entité immatérielle, c’est-à-dire à l’âme, pour caractériser l’intelligence. Aussi, positivement, l’intelligence est-elle caractérisée par Turing comme la manifestation *verbale* d’un comportement rationnel au sein d’une interaction prenant la forme d’un dialogue.

On infère donc de l’existence de manifestations verbales la présence d’une capacité intelligente dont la caractérisation n’est pas poursuivie plus avant : l’intelligence joue donc le rôle d’une “boîte noire” dont on analyse seulement les manifestations. Identifier au sein de boîtes noires un comportement intelligent sur la fois de traces verbales permet d’induire une commune appartenance de celles-ci au genre “entité intelligente”. De cette inférence, on déduit la possibilité d’une “intelligence artificielle”, définie comme le comportement qui aurait toutes les apparences d’un comportement intelligent. L’intelligence ayant été caractérisée précisément par cette apparence extérieure, il n’est pas besoin de supposer la nécessité d’une quelconque intériorité supplémentaire pour parvenir à la définir.

12. La différence physique entre les êtres humains et les ordinateurs digitaux

La caractérisation de l’intelligence telle que Turing cherche à l’établir en fait une notion tout abstraite : les traces verbales nécessaires pour caractériser les boîtes noires font que la caractérisation de ces dernières ne dépend pas d’un substrat matériel particulier. Mais ni les êtres humains ni les ordinateurs digitaux ne sont, à première vue, des boîtes noires. Il s’agit au contraire dans les deux cas d’entités matérielles, les unes possédant des corps au sens biologique du terme et les autres des composants électriques et électroniques[180]. Tenter de montrer que l’identification entre les êtres humains réels d’une part et les incarnations de la machine universelle que sont les ordinateurs digitaux de l’autre se vérifie même au sein du monde physique implique donc l’on parvienne *à mettre de côté* ce qui fait les différences physiques entre les deux types d’entités. Comment réussir à les constituer en boîte noire ?

Pour réussir à atteindre ce but, il faut réussir à trouver un niveau de description adéquat au sein duquel les différences physiques perceptibles entre les êtres humains et les ordinateurs digitaux soient sans pertinence. Ce niveau de description est précisément celui de l’étude de l’intelligence définie comme comportement essentiellement verbal dans une partie de jeu de l’imitation. Le jeu de l’imitation permet donc de renvoyer dos à dos deux attitudes opposées mais qui sont en fait complémentaires : mettre l’accent sur l’aspect *physique* des entités étudiées conduit à présupposer l’existence d’une caractéristique *immatérielle* pour rendre compte de la spécificité du comportement *rationnel* dans le cas strictement *humain*. Ainsi l’immatérialité de l’intelligence ne serait-elle paradoxalement que la traduction abstraite d’une différence physique entre l’être humain et les autres entités présentes dans le monde matériel. En éliminant la référence à l’aspect physique des entités étudiées, Turing n’a plus besoin de présupposer un aspect

[180] Je laisse de côté le fait qu’une machine de Turing puisse s’incarner dans des substrats physiques très divers, que ce soit du métal ou même de la lumière. Le point à souligner est qu’il faut qu’elle possède *un* substrat physique quel qu’il soit.

immatériel à la notion d'intelligence mais seulement un aspect *abstrait* et peut élargir de ce fait le champ d'application de la notion à d'autres entités que l'être humain. De ce point de vue, on peut dire que l'intelligence n'est *ni corps ni âme mais seulement concept*.

Mais peut-on éliminer entièrement tout rapport à l'aspect physique des entités étudiées ? Tout comportement ne se donne-t-il pas à voir physiquement ? Si bien sûr : toute manifestation physique d'un comportement doit nécessairement s'exprimer dans l'univers matériel; mais si l'on se limite, au sein de l'univers physique, à l'aspect verbal des comportements, mieux : à la trace écrite de la parole et mieux encore : à la trace *dactylographiée* de celle-ci, on dissocie entièrement la manifestation du comportement des particularités physiques de l'entité qui en est l'origine. Ainsi est-ce en suivant un enchaînement de traces qui vont de l'écriture dactylographiée à la parole et de celle-ci au comportement rationnel supposé en être la source que l'on peut attribuer, à toute entité physique quelle qu'elle soit, le caractère non-verbal de l'intelligence sans avoir à présupposer l'immatérialité de cette dernière. Cet enchaînement de traces permet d'inférer d'une manifestation bi-dimensionnelle de l'intelligence (l'écriture dactylographiée[181]) une manifestation tri-dimensionnelle de celle-ci (la parole[182]) sans faire référence aux particularités tridimensionnelles que toute entité physique possède dans l'espace. C'est cette articulation tout à fait particulière des dimensions deux et trois par le biais d'un jeu qui rend possible la constitution d'une psychologie spécifique dont l'unique objet d'étude est la manifestation verbalisée de l'intelligence[183].

Quelle est la méthode utilisée par cette psychologie ? Si les êtres humains et les ordinateurs digitaux devenaient indiscernables au sein du jeu de l'imitation, alors le modèle de la machine de Turing serait suffisant pour étudier les

---

[181] Schématiquement, l'écriture se donne en effet en deux dimensions : l'une pour le tracé des caractères composés de segments de droite et l'autre pour le temps qui permet de passer d'un caractère à l'autre.

[182] La parole en tant qu'elle est un système d'ondes dans l'espace est de nature tri-dimensionnelle.

[183] Du point de vue de l'histoire de la logique, on peut interpréter le jeu de l'imitation comme la résurgence de la problématique de l'ancienne logique qui, d'Aristote à Port-Royal, s'est constituée en théorie du *discours*, au sein de la dernière-née des branches de la logique mathématique dans les années 30 de ce siècle : la théorie de la *calculabilité*.

manifestations de l’intelligence quelles qu’elles soient et constituer de ce fait la méthode nécessaire à une investigation psychologique d’un nouveau type. Cette psychologie suppose donc seulement de faire l’hypothèse, dans toute entité physique, être humain ou ordinateur digital, de la présence *cachée mais nécessaire* d’un fonctionnement “mécanique” identique à celui d’une machine de Turing. On peut alors caractériser ce que Turing entend par “entité intelligente” : est qualifiée d’intelligente toute boîte noire que l’on suppose physiquement tridimensionnelle et qui susceptible de manifester un comportement verbal analysable en deux dimensions par le biais du modèle de la machine de Turing.

Si l’on pouvait effectivement s’en tenir à cet enchaînement de “traces” au sein du jeu de l’imitation, alors il deviendrait possible de fonder la différence entre le concept abstrait d’intelligence et les entités physiques qui lui servent de support. Les conséquences pour une théorie de l’esprit seraient immenses : toutes les manifestations de l’intelligence seraient considérées comme faisant partie d’un domaine autonome de nature abstraite qu’il serait possible d’étudier scientifiquement pour lui-même. En adoptant une autre attitude à l’égard de la réalité physique, Turing compte donc montrer que l’intelligence ne doit plus être considérée comme une notion immatérielle mais bien plutôt comme un concept abstrait. Ce domaine abstrait serait précisément celui de l’intelligence artificielle, situé en quelque sorte “à cheval” sur le domaine humain et celui des machines. C’est donc la façon originale dont Turing parvient à distinguer réalité physique et concept abstrait qui constitue le nerf de son argument.

2. Exposition des règles du jeu

21. Traduction des deux premières sections de “Computing Machinery and Intelligence”

«1. Le jeu de l’imitation
Je propose de prendre en considération la question, “Les machines peuvent-elles penser ?”. Il faudrait commencer, pour ce faire, par la définition des termes de “machine” et de “penser”. Les définitions pourraient être construites pour faire en sorte qu’elles correspondent autant que possible à l’usage normal des mots, mais cette attitude serait dangereuse. S’il fallait, pour trouver le sens des mots “machine” et “penser”, examiner comment ils sont utilisés couramment, on n’échapperait pas à la conclusion que l’on doit

chercher le sens et la réponse à la question “Les machines peuvent-elles penser ?” dans une étude statistique tel qu’un sondage d’opinion. Mais c’est absurde. Plutôt que de me risquer à une telle définition, je vais remplacer la question par une autre, qui s’y rapporte de près et que l’on peut exprimer en des termes relativement peu ambigus.

La nouvelle formulation du problème s’apparente à un jeu que nous appellerons le “jeu de l’imitation”. Il se joue à trois, un homme (A), une femme (B) et un interrogateur (C), qui peut être de l’un ou l’autre sexe. L’interrogateur demeure dans une pièce différente de celle de deux autres joueurs. Le but du jeu, pour l’interrogateur, est de déterminer lequel des deux est l’homme et lequel la femme. Il les connaît sous les labels X et Y et à la fin du jeu, il doit dire soit “X est A et Y est B” soit “X est B et Y est A”. L’interrogateur a le droit de poser à A et B des questions telles que :

C: X pourrait-il ou pourrait-elle, s’il vous plaît, me dire la longueur de ses cheveux ?
Supposons que X est vraiment A et qu’il lui faut donner une réponse. Le but de A dans le jeu est de tenter d’induire C en erreur. Sa réponse pourrait donc être :

“«Mes cheveux sont coupés à la garçonne et les mèches les plus longues font à peu près vingt centimètres.

Pour faire en sorte que les tons de voix ne viennent pas en aide à l’interrogateur, les réponses devraient être écrites, ou mieux encore, dactylographiées. La configuration idéale serait de disposer d’une téléimprimante communiquant à travers les deux pièces. On peut aussi concevoir que questions et réponses soient répétées par un intermédiaire. Le but du jeu pour le troisième joueur (B) est de venir en aide à l’interrogateur. La meilleure stratégie pour celle-ci est sans doute de donner des réponses vraies. Elle peut ajouter des remarques à ses réponses comme “je suis la femme, ne l’écoutez pas !” , mais cela n’aboutirait à rien, car l’homme peut faire des remarques semblables.

Nous posons maintenant la question : “Que se passera-t-il si l’on substitue une machine à A dans le jeu ?” L’interrogateur se trompera-t-il autant de fois quand le jeu est joué ainsi que lorsqu’il est joué entre un homme et une femme ? Ces questions remplacent la question originelle, “Les machines peuvent-elles penser ?”

2. Critique du nouveau problème

De même que l’on pose la question : “Quelle est la réponse à cette nouvelle formulation de la question ?”, on peut se demander : “La dernière question mérite-t-elle qu’on s’interroge sur elle ?”. Nous tenterons de mener une enquête concernant cette dernière question sans autre forme de procès, coupant ainsi court à une régression à l’infini.

Le nouveau problème a l’avantage de tracer une ligne assez nette entre les capacités physiques et intellectuelles d’un homme. Aucun ingénieur ni aucun chimiste ne prétend être capable de produire un matériau indiscernable de la peau humaine. Il est possible que ceci soit réalisé dans l’avenir mais cette invention serait-elle disponible que nous devrions sentir combien rendre plus humaine une “machine pensante” en l’habillant de cette chair artificielle serait hors de propos. La formulation que nous avons donnée au problème se fait sentir dans la condition qui empêche l’interrogateur de voir, de toucher ou d’entendre les voix des autres adversaires. On peut montrer les autres avantages du critère proposé, en fournissant certaines questions et réponses spécifiques :

Question : Ecrivez-moi, s’il vous plaît, un sonnet au sujet du pont sur le Forth.
Réponse : Ne comptez pas sur moi pour ça. Je n’ai jamais été capable d’écrire des poèmes.
Q :Ajoutez 34 957 et 70 764.
R: (En silence pendant 30 secondes puis donne la réponse) 105 621.
Q Jouez-vous aux échecs ?
R: Oui.
Q: J’ai mon roi en C8 et plus d’autres pièces. Il vous reste seulement votre roi en C6 et une tour en A1. A vous de jouer. Que jouez-vous ?
R : La tour en A8 , échec et mat.

La méthode des questions et des réponses semble apte à introduire n’importe quel domaine de l’activité humaine que nous souhaiterions inclure. Nous ne souhaitons pas pénaliser la machine par son incapacité à briller dans les concours de beauté, ni pénaliser l’homme parce qu’il perdrait à la course contre un avion. Les conditions de notre jeu ôtent toute pertinence à ces incapacités. Les “témoins” peuvent se vanter autant qu’ils leur plaît de leur charme, de leur force ou de leur héroïsme, mais l’interrogateur ne peut pas exiger de démonstrations pratiques.

On peut critiquer le jeu en faisant remarquer que la machine est trop lourdement

handicapée. Si l'homme essayait de faire croire qu'il est la machine, il ferait très certainement mauvaise figure. On le reconnaîtrait immédiatement par sa lenteur et son inexactitude en arithmétique. Est-ce que les machines ne seraient pas en mesure d'effectuer quelque chose qui devrait être décrit comme de la pensée mais qui serait très différent de ce qu'un homme effectue ? Cette objection est très forte mais au moins pouvons-nous dire que, quoi qu'il en soit, si une machine est construite de telle sorte qu'elle soit en mesure de jouer de façon satisfaisante au jeu de l'imitation, nous n'avons pas à nous inquiéter de cette objection.

On pourrait arguer du fait que lors d'une partie de "jeu de l'imitation", la meilleure stratégie pour la machine pourrait être différente de l'imitation du comportement d'un homme. C'est peut-être le cas, mais je ne crois pas qu'il y aurait des effets de conséquence de ce type. Quoi qu'il en soit, je n'ai pas l'intention de faire des recherches sur la théorie du jeu et je ferai l'hypothèse que la meilleure stratégie pour la machine est de fournir des réponses qu'aurait naturellement fournies un homme.»

22. Les règles du jeu

Revenons un instant sur les règles du jeu. Il se joue à trois : un homme (appelé A), une femme (B) et un interrogateur dont le sexe n'est pas précisé (C). L'interrogateur est séparé de l'homme et de la femme (il se trouve, par exemple, dans une autre pièce). Le but du jeu pour l'interrogateur est de déterminer, des deux personnes qui se trouvent hors de sa vue, quel est l'homme et quelle est la femme. Idéalement, la communication entre les joueurs ne doit pas se faire par l'intermédiaire de la parole mais au moyen d'une machine à écrire et sans contact physique direct, pour limiter le plus possible toute interférence liée à l'apparence physique des joueurs et ne faire entrer en jeu que les manifestations écrites de leurs facultés intellectuelles. Chaque joueur peut induire l'interrogateur en erreur si cela lui paraît être la meilleure stratégie pour cacher son identité : dans ce cas, il lui faut *imiter* les réponses qu'il suppose être celles que donnerait l'autre joueur, de sexe opposé. D'où le nom du jeu. Le jeu est terminé quand l'interrogateur arrive à fournir la bonne réponse. On appellera cette formule du jeu le "jeu n°1".

Turing fait ensuite subir aux règles décrites une transformation : dans la nouvelle formule du jeu (appelé par nous "jeu n°2"), le joueur masculin A sera remplacé, à l'insu de l'interrogateur, par un ordinateur digital qui sera programmé de telle sorte qu'il imite le type de réponse du joueur masculin qu'il remplace.

La question que pose Turing est la suivante : l'interrogateur peut-il se rendre compte par lui-même que l'on est passé du jeu n°1 au jeu n°2, c'est-à-dire qu'il n'a plus affaire à deux êtres humains de sexe opposé mais à une femme d'une part et à un ordinateur d'autre part ? La réponse de Turing est "non" : l'interrogateur n'a pas les moyens de faire la distinction entre les joueurs dans la deuxième formule du jeu. Mais les moyens mis en œuvre par Turing pour aboutir à cette réponse négative demandent à être explicités.

23. Le critère de la différence des sexes

Remarquons tout d'abord que la règle du jeu porte sur la détermination du sexe des joueurs et que c'est précisément cette différence qui doit, au cours d'une partie, c'est-à-dire dans la transformation du jeu n°1 en jeu n°2, perdre tout objet et ne plus constituer qu'une illusion de la part de l'interrogateur. Du point de vue du lecteur observant en pensée une partie, la différence proprement sexuelle entre homme et femme doit être remplacée par une différence non sexuelle entre femme et ordinateur. C'est cette illusion touchant à la différence des sexes qui doit quitter le lecteur de l'article quand il observe, en pensée, le déroulement complet d'une partie. Se garder de cette illusion constitue donc, pour un observateur extérieur, le critère même de réussite du jeu. Une fois en possession de ce critère, il devient possible d'interpréter les performances de l'interrogateur et des joueurs comme celles d'un système organisé. La différence des sexes est-il nécessaire à la bonne marche du jeu ? Peut-on imaginer un jeu de l'imitation dans lequel un autre critère serait utilisé ?

En choisissant la question de la différence des sexes, Turing a choisi, du point de vue physique, la différence physique la plus originaire chez les êtres humains puisque la différence sexuelle est à la source des différences physiques entre l'homme et la femme. Or c'est bien cela qu'il s'agit de prouver, dans la première formule du jeu d'abord et dans la seconde ensuite : montrer que l'on peut éliminer l'aspect physique des joueurs en établissant définitivement la différence entre le physique et l'intellectuel et réussir de ce fait à montrer que les opérations intellectuelles peuvent s'incarner dans les substrats les plus divers. La transformation du jeu n°1 en jeu n°2 consiste donc à *déplacer* la différence illusoire entre divers aspects physiques (homme et femme ou encore ordinateur et femme) vers la différence réelle entre le physique pris globalement et l'intellectuel pris globalement (substrats physiques, intelligence).

Pour montrer au lecteur extérieur à la partie l'universalité du concept d'intelligence, deux étapes sont donc nécessaires. Premièrement, il faut réussir à montrer que les différences physiques entre l'homme et la femme, différences qui reposent en dernière instance sur la différence des sexes, ne comptent pas : c'est le but du jeu n°1. Deuxièmement, il faut montrer que les différences physiques entre la femme d'une part et un ordinateur digital convenablement programmé d'autre part, ne comptent pas non plus : c'est le but du jeu n°2. Cette deuxième formule du jeu paraît, au premier abord, curieuse : pourquoi avoir choisi le cas de la femme et non celui de l'homme dans la "compétition" avec l'ordinateur digital ?

On pourrait objecter que cette question n'a pas de sens et que le choix de la femme est indifférent par rapport à l'issue du jeu. Dans cette optique, on dirait que Turing aurait pu aussi bien choisir le cas de l'homme : femme et homme seraient, de ce point de vue, complètement interchangeables au sein de la deuxième formule du jeu. Mais cette interprétation n'est pas soutenable car elle revient à *faire l'hypothèse que la différence entre homme et femme est, a priori, déjà supprimée* et que l'issue du jeu n°1 est déjà acquise. Or rien n'est encore certain : on ne peut pas présupposer l'issue du jeu n°1 avant de s'être assuré, par le jeu, de l'aspect interchangeable de la femme et de l'homme. De plus, en privilégiant de façon *a priori* une issue "positive" à la partie, à savoir la disparition de la différence des sexes, le jeu de l'imitation se réduirait finalement au jeu n°2, qui oppose au départ l'être humain quel que soit son sexe et l'ordinateur puis qui rend possible une réduction de opposition[184]. Il faut donc interpréter autrement le remplacement de l'homme par un ordinateur et la "compétition" que se livrent la femme et l'ordinateur.

Si l'on accorde qu'il y a, au sein du jeu n°1, une certaine nécessité à avoir choisi la notion de différence des sexes pour opposer le plus radicalement possible d'un point de vue physique les deux joueurs, homme et femme, on peut se laisser guider, dans le cas du jeu n°2, par le même raisonnement : de même que la femme est ce qu'il y a de plus opposé à l'homme au sein du genre humain, de même la femme serait ce qu'il y a de plus opposé à l'ordinateur digital quand on fait une comparaison entre l'humain et le non-humain. Ainsi la femme jouerait-elle toujours, dans le jeu de l'imitation, le rôle de *terme le plus opposé* : à l'homme dans le cas d'une différence interne au genre humain et à l'ordinateur dans le cas

[184] C'est d'ailleurs ainsi que l'on résume habituellement le jeu de l'imitation.

d'une différence entre le genre humain et le genre non-humain, "mécanique". Quelle conclusion en tirer du point de vue de ce qu'il s'agit d'établir, à savoir l'universalité du concept d'intelligence ?

Au vu des remarques que nous venons de faire, la réponse à la question de l'universalité potentielle de l'intelligence engage une deuxième question : l'ordinateur remplace-t-il l'homme ou le genre humain dans le jeu n°2 ? La réponse à cette dernière *exige que l'on puisse établir quelle est l'issue du jeu n°1*, car c'est elle qui peut finalement permettre le passage à la deuxième formule du jeu. Deux cas sont envisageables.

S'il est vrai que l'issue du jeu n°1 est nécessairement positive, c'est-à-dire que l'interrogateur ne parvient pas à établir de distinction entre homme et femme, il devient possible d'opposer le genre humain pris globalement et le non-humain mécanique. Si, ensuite, l'interrogateur n'aperçoit pas de changement de règle entre le jeu n°1 et le jeu n°2, on aura par conséquent dépassé le cas particulier de l'espèce humaine et atteint le but que l'on s'est fixé : faire de la notion d'intelligence un *concept* abstrait en séparant radicalement les manifestations de l'intelligence de ses substrats physiques. La notion d'intelligence sera alors considérée à bon droit comme un concept et ses incarnations dans l'espèce humaine ou dans les ordinateurs digitaux n'apparaîtront plus dès lors que comme des espèces différentes d'un même genre, susceptible d'être décrit par le concept de machine de Turing.

Dans le cas où il serait impossible d'établir la nécessité d'une issue positive au jeu n°1, alors c'est la possibilité d'une extension de la portée de la notion d'intelligence qui serait remise en question. Il faudrait alors revenir à l'analogie établie par Turing dans les deux formules du jeu : la femme est à l'homme ce que la femme est à l'ordinateur digital et tenter de voir en quoi la question de la différence des sexes change l'extension que l'on donne à la notion d'intelligence. Dans cette optique, deux questions sont à considérer. D'une part, il faut se poser la question de la place du terme extrême que représente la femme, puisque, au sein de l'ensemble des deux joueurs, c'est le seul terme qui ne change pas dans les deux formules du jeu. D'autre part, il faut comprendre comment s’effectue le remplacement de l’homme (en tant qu’il s’oppose à la femme et non en tant que représentant du genre humain) par l’ordinateur digital du point de vue des rapports qu’entretiennent la notion de pensée et celle de machine.

Il nous faut donc étudier les différentes interprétations que l’on a donné du jeu pour savoir si c’est le premier ou le deuxième cas de figure qu’il faut retenir.

Chapitre II

Interprétation formaliste du jeu de l'imitation

L'interprétation la plus courante de la démarche de Turing dans "Computing Machinery and Intelligence" est une interprétation formaliste ou tout au moins formalisante. Elle consiste à interpréter la démarche de celui-ci en présupposant la validité de la thèse de Turing, c'est-à-dire le passage toujours possible de la notion informelle d'algorithme à la notion formelle de machine.

Cette interprétation est la plus couramment admise parce qu'elle rend directement compte du projet de l'intelligence artificielle : la thèse de Turing dépasserait le cadre logico-mathématique de la formalisation de la notion informelle d'algorithme puisqu'elle s'appliquerait au cas de la pensée informelle en général, comme le montrerait le jeu de l'imitation. La conséquence, fidèle au mouvement de la thèse de Turing, est que la notion de machine aurait aussi une extension que l'on ne lui soupçonnait pas, puisqu'elle pourrait s'appliquer au cas de l'intelligence en général.

Remarquons que l'utilisation de la thèse de Turing pour corroborer le projet de l'intelligence artificielle est beaucoup moins sujette à controverse que l'utilisation couramment faite de théorèmes de logique mathématique en vue du même but, comme par exemple l'utilisation des théorèmes de Gödel[185]. Il est en

---

[185] On sait que les théorèmes de Gödel ont été utilisés dans des argumentations anti- puis pro-mécanistes. Cf. pour la position anti-mécaniste : **J. R. Lucas**, "Minds, Machines and Gödel", Philosophy, XXXVI, 1961; trad. franç. dans [*Pensée et Machine*, Champ Vallon, Seyssel, 1983],

effet difficile - voire impossible - d'appliquer une argumentation gödelienne portant sur la notion précise de système formel à une notion aussi peu formelle que "l'esprit" ou "l'intelligence"[186]. En revanche, la thèse de Turing, dans la mesure même où elle décrit le passage de l'informel au formel, se prête beaucoup mieux à une argumentation de ce type. Il y a donc *a priori* plus de raisons de penser que c'est sur le terrain d'une "extension" possible de la portée de la thèse de Turing que l'on peut corroborer le point de vue mécaniste de l'intelligence artificielle que sur le terrain directement formel.

A quoi tient l'aspect formaliste ou pour mieux dire, "formalisant", de l'interprétation du jeu de l'imitation ?

1. Parenté du jeu de l'imitation avec la perspective formelle

A la lecture de l'article de Turing, il apparaît clairement que l'itinéraire décrit par le jeu consiste à éliminer progressivement les réticences du lecteur à l'égard d'une extension de la notion d'algorithme à des classes de phénomènes qui semblent échapper à cette notion : ce qui apparaît au début de l'article comme "bien connu" chez le lecteur - la différence entre un être humain et un ordinateur - , doit finir par disparaître au moyen du processus mis en place par le jeu.

Les buts que Turing chercheraient à atteindre par le truchement de l'expérience du jeu de l'imitation seraient donc les suivants. Du point de vue le plus particulier, l'article chercherait à montrer que les deux formules du jeu sont identiques et qu'il est possible de ce fait d'étudier par le biais du même concept d'intelligence les manifestations dactylographiées d'entités quelconques au sein du jeu. Ce résultat n'a pas d'intérêt en lui-même. Il corroborerait un autre but de caractère plus général, à savoir le fait que les manifestations dactylographiées sont le signe de la présence d'intelligence dans des entités quelconques. Ce résultat viserait un résultat plus général encore : la validation du modèle discret de la

---

pp. 81-97, et pour la position pro-mécaniste : **J. C. Webb**, *Mechanism, Mentalism and Metamathematics : an Essay on Finitism*, Reidel, Dordrecht, 1980, p. 336 sq.

[186] C'est ce que remarque J. Bouveresse quand il dit que l'être humain est un système informel auquel appliquer la notion formelle de consistance (indispensable à la démonstration du théorème de limitation des formalismes) paraît bien étrange. Cf. **J. Bouveresse**, *La parole malheureuse*, Editions de Minuit, Paris, 1971, p. 406.

machine de Turing pour l'étude des comportements intelligents quels qu'ils soient. L'extension du concept d'intelligence, nécessaire pour atteindre ces buts, s'opère par un mouvement en deux étapes.

Premièrement, en se situant à l'extérieur d'une partie et en se mettant à la place de l'interrogateur, le lecteur doit finir par s'apercevoir que l'interrogateur ne doit pas parvenir à faire la différence entre l'homme et la femme (ni, une fois la formule du jeu changée à son insu, entre cette différence elle-même et celle qui est censée exister entre l'être humain et le non-humain). Réciproquement, en se mettant à la place des joueurs, le lecteur doit finir par comprendre qu'il doit leur être possible de tromper l'interrogateur sur leur sexe ou leur humanité. Un concept central utilisé dans "On Computable Numbers …" a peut-être contribué, comme nous allons le voir, à l'interprétation formaliste de ce premier mouvement : le concept de décision.

Deuxièmement, en considérant seulement les messages dactylographiés qui sont échangés, le lecteur comprend qu'ils doivent être suffisants pour rendre compte des manifestations d'intelligence. Il ne s'agit donc plus ici de se placer du point de vue de la pensée humaine, mais bien de dépasser ce cadre et de considérer l'intelligence en général comme un concept explicable par le modèle de la machine de Turing. Bref, le lecteur doit adopter à l'égard des messages dactylographiés l'attitude d'une machine de Turing : le remplacement de l'homme par une machine de Turing dans le jeu n°2 vise en fait le remplacement chez le lecteur d'une attitude "humaine" à l'égard du jeu par une attitude "mécanique". Ainsi le remplacement de l'homme est-il concomitant du remplacement du *lecteur* par une machine : c'est ce dernier remplacement, effectué à l'insu du lecteur lui-même - Turing n'en parle d'ailleurs à aucun moment, le texte de l'article devant opérer par lui-même cette modification -, qui marque la fin de l'illusion sur la pertinence de la différence des sexes, c'est-à-dire de l'humanité, pour la caractérisation de l'intelligence[187]. Quand on se situe dans le cadre de

---

[187] On peut donc dire que la différence corps / âme que Turing récuse pour définir l'intelligence vise à être remplacée par une différence intellectuelle entre forme et contenu : de même que mettre l'accent sur l'aspect physique de l'homme impliquait, quand il s'agissait de rendre compte de son comportement rationnel, de postuler l'existence d'une forme immatérielle, l'âme, de même ici, le

l'interprétation formaliste, il y a donc en fait, à l'issue d'une partie de jeu n°2, *deux* machines de Turing présentes dans le jeu : la première remplace l'homme à l'insu de l'interrogateur et la seconde remplace le lecteur à son propre insu. Ces machines sont-elles de type machine-*a*, machine-*c*, machine-*o* ou d'un autre type de machine non encore précisé ? C'est la notion *d'imitation*, héritée de "On Computable Numbers …", qui permettra de répondre à cette question.

Il nous faut donc étudier, dans le jeu, les deux étapes de la constitution de l'universalité de la notion d'intelligence et leur articulation.

11. Première étape de la constitution du concept universel d'intelligence : la question de la décision

Quel type de stratégie doivent employer l'interrogateur et les joueurs au cours de la première étape du jeu ? Est-ce une stratégie d'omniscience ou une stratégie pragmatique qui est le mieux adaptée à la situation d'une partie ?

Pour réussir à mener à bien une stratégie d'omniscience, il faudrait être capable de faire une analyse mathématique de toutes les stratégies possibles de tous les participants. Mais le jeu, en exigeant seulement que les échanges entre participants prennent la forme d'un dialogue orienté vers la détermination des sexes des joueurs, instaure une règle si peu contraignante qu'il paraît difficile de concevoir comment on pourrait en faire une analyse en termes mathématiques. Dès lors, il semble nécessaire à l'interrogateur et aux joueurs d'adopter une stratégie pragmatique qui, en tant que telle, fait référence à l'existence d'une psychologie réelle chez l'adversaire. C'est dans ce contexte que l'interrogateur doit prendre une *décision* quant au sexe des joueurs.

Par une extension métaphorique du concept de décision appliquée au cas informel de l'intelligence humaine, on voit que ce que l'interrogateur est censé *ne pas* parvenir à faire, c'est précisément à prendre une décision qui consisterait à répondre par oui ou par non à la question de la détermination du sexe des joueurs dans la première formule du jeu ou de leur caractère humain ou non-humain dans

---

remplacement de l'intelligence de l'homme par un ordinateur digital dans le jeu n°2, implique de postuler l'existence d'une machine qui joue le rôle de forme immatérielle par rapport à des contenus immatériels, les joueurs humain ou non-humain et qui prend la place du lecteur.

la seconde. Il existe donc une parenté entre la problématique générale de la décision dans les systèmes formels et le jeu de l'imitation. Mais, comme on va le voir, il s'agit d'une parenté superficielle car le jeu de l'imitation se distingue de la perspective formelle sur un point capital : son argumentation ne relève pas directement de la logique.

Si l'on rapporte la perspective formelle et plus précisément la question de la décision à la façon dont Turing l'a élaborée dans "On Computable Numbers …", on voit que la parenté entre les deux domaines est finalement très limitée. Dans le cas de la perspective formelle en effet, on démontre logiquement qu'il n'y pas de machine qui résoudrait le problème de l'arrêt pour toute machine, puisque l'on se heurte, en faisant cette supposition, à une contradiction. Dans le jeu de l'imitation, l'interrogateur ne parvient pas à une décision et en cela la perspective semble bien être la même que celle du problème de l'arrêt. Mais cette absence de décision *n'est pas le fait d'une argumentation logique* : elle repose sur la constatation que si l'interrogateur ne parvient pas à prendre une décision, alors le jeu doit se poursuivre *indéfiniment*, à moins d'être arrêté par une décision extérieure. La réussite de ce que Turing cherche à montrer par le biais du jeu exige donc l'échec de l'interrogateur et son absence de décision, décision qui est reléguée à l'extérieur du jeu, par exemple en supposant l'existence d'un arbitre qui, au bout d'une certaine durée de jeu, décidera d'interrompre la partie et déclarera l'échec de l'interrogateur.

L'argumentation logique reposant sur l'usage du principe de non-contradiction est donc remplacée, dans le contexte du jeu de l'imitation, par une argumentation temporelle qui n'est pas autre chose que l'affirmation de l'illimitation temporelle du jeu. Cette illimitation temporelle exigerait alors l'intervention extérieure d'un arbitre. L'usage du principe de non-contradiction serait remplacé par la différence de point de vue entre l'intérieur et l'extérieur du jeu. Aussi, dans le cas du problème de l'arrêt, la démonstration prouve-t-elle de façon *absolue* l'absence d'une machine qui résoudrait universellement la question de l'arrêt des machines pour toute entrée donnée, tandis que dans le jeu, l'absence d'une intelligence susceptible de prendre une décision concernant le sexe des joueurs n'est-elle que *relative* au point de vue sur le jeu : soit à l'intérieur de celui-ci, soit à l'extérieur. L'absence d'arrêt - de la machine et d'une partie - n'a donc pas le même sens dans les deux perspectives.

Pourquoi la perspective logique n'a-t-elle pas été poursuivie de façon similaire dans le jeu de l'imitation ? Parce que la décision que doit prendre l'interrogateur ne relève pas du domaine *déjà formel* touchant la question de l'arrêt des machines de Turing mais de la différence à faire entre les aspects intellectuels et physiques des joueurs. Or ceci n'est pas un problème relevant du domaine formel mais un problème qui relève du *passage* du domaine informel au domaine formel. En effet, on peut dire que l'expérience du jeu de l'imitation a réussi si l'interrogateur ne parvient pas à prendre de décision concernant le sexe des joueurs, c'est-à-dire s'il ne parvient pas à séparer l'intellectuel du physique : s'il y parvenait, il pourrait établir en quoi l'aspect masculin des réponses de l'homme et l'aspect féminin des réponses de la femme se reflètent dans leur discours.

La différence profonde entre les deux articles se manifeste donc directement dans leur différence de perspective temporelle. Dans le cas de "On Computable Numbers …", on a vu que le calcul indéfiniment long des décimales des réels calculables ne contrariait en rien la définition précise de l'acte de calcul. Dans le jeu de l'imitation de "Computing Machinery and Intelligence" au contraire, l'aspect indéfini de la durée d'une partie est l'indice du fait qu'une perspective non-logique est mise en place, celle qui consiste à essayer de distinguer l'intellectuel du physique : l'interrogateur cherche indéfiniment à atteindre ce but, alors que cette séparation est déjà acquise dans la perspective formelle de "On Computable Numbers …" puisque dès que l'on a trouvé la table d'instructions d'une machine capable de calculer le développement décimal d'un nombre réel, on se trouve dans le registre du formel.

Cette différence de perspective indique une première limite au rapprochement entre les deux articles effectué d'un point de vue formaliste.

12. Deuxième étape de la constitution du concept universel d'intelligence : la

question de l’imitation

Dans la deuxième étape de la constitution de la notion d’intelligence en concept universel, celle qui consiste à ne prendre en considération que les messages dactylographiés rédigés par les participants au jeu, on postule une identité entre l’imitation mécanique et l’imitation ludique. C’est en effet la notion d’imitation qui permet de construire une analogie entre le formalisme de la machine de Turing et le jeu de l’imitation : il y a le même rapport entre une machine universelle de Turing et une machine de Turing particulière qu’entre la définition générale de l’intelligence et un type particulier d’intelligence (comme l’intelligence humaine). Plus précisément, l’interprétation formaliste établit une analogie entre la façon de programmer une machine de Turing et la façon de poser des questions au jeu de l’imitation.

On peut en effet interpréter les dialogues du jeu de l’imitation comme les manifestations d’une relation entrée / sortie : l’interrogateur inscrit des données en entrée (ses questions) et les joueurs émettent des réponses (en sortie). Le jeu de l’imitation constitue alors une boite noire, qui, en tant que telle, est modélisable par le biais du modèle de la machine de Turing[188] : cette interprétation justifie donc le projet de l’intelligence artificielle qui voit dans le modèle de la machine de Turing le moyen de caractériser l’intelligence. Dans cette analogie, l’interrogateur est semblable à un programmeur qui, au départ, introduit en entrée les données à traiter par la machine. Les joueurs font office de table d’instructions pour la machine en question : de même que la table d’instructions est constituée par le couple (symbole lu; état interne) de même, à la lecture des questions reçues, chaque joueur réagit selon son propre état interne et propose en sortie une réponse, c’est-à-dire un résultat. L’interrogateur-programmeur prend alors une décision concernant la vérité ou la fausseté des deux résultats proposés et introduit de nouvelles données en entrée sous forme d’une question.

La décision que doit prendre l’interrogateur-programmeur revient à faire un choix entre deux résultats. Ce choix doit-il être motivé ou peut-il être aléatoire

---

[188] Cette interprétation du jeu de l’imitation est décrite par Pierre Wagner dans **P. Wagner**, *Machine et pensée : l’importance philosophique de l’informatique et de l’intelligence artificielle*, thèse de doctorat de philosophie de l’Université Paris-I, janvier 1994, pp. 42-43.

? C'est la nature de la stratégie à employer par l'interrogateur-programmeur, stratégie d'omniscience ou stratégie pragmatique, qui fait ici question. La stratégie employée ne peut plus être une stratégie pragmatique, puisque les manifestations d'intelligence, quelles qu'elles soient, sont censées être toutes explicables sans faire référence à une psychologie humaine spécifique. C'est la stratégie d'omniscience qui doit permettre de rendre compte des bluffs que pourraient formuler l'interrogateur et les joueurs.

L'analogie formaliste entre le jeu de l'imitation et le mode de fonctionnement d'une machine de Turing, analogie qui constitue la deuxième étape de la constitution de l'intelligence en concept universel, vaut-elle à la fois pour le premier et le deuxième jeu ?

Dans le cas du premier jeu, la question est de savoir si la différence des sexes a bien été écartée. On a vu que celle-ci ne pouvait être écartée que si la décision concernant la fin de la partie était repoussée à l'extérieur du jeu sur la personne d'un arbitre potentiel. Qu'en est-il alors de cet arbitre et de sa prise de décision ?

Comme on l'a déjà remarqué, la nécessité de postuler un arbitre extérieur au jeu qui prenne la décision d'arrêter la partie et de déclarer l'échec de l'interrogateur pour le cas où celui-ci ne parviendrait pas à distinguer l'homme de la femme ou la femme de l'ordinateur exige finalement que ce point de vue extérieur soit, à la fin d'une partie de jeu n°2, occupé par une machine. En effet, ce que vise le jeu de l'imitation, c'est bien à convaincre le lecteur qui observe de l'extérieur une partie que la différence entre être humain et ordinateur du point de vue de la notion d'intelligence est illusoire et qu'il s'agit de dépasser cette opposition. Or comme le lecteur occupe la position d'arbitre à l'extérieur du jeu, la disparition de cette illusion consiste précisément à ce que le jugement du lecteur ne se distingue plus de celui d'une machine. Deux machines doivent donc être prises en considération dans le deuxième étape de la constitution de la notion d'intelligence en concept universel : la machine *interne* au jeu qui remplace l'homme et la machine *externe* au jeu qui doit finir par remplacer le lecteur. Comment caractériser ces machines ?

Selon l'analogie formaliste, c'est en se référant à la typologie des machines de Turing qu'il devient possible de caractériser avec précision les machines du jeu. Plus précisément, on peut réussir à décrire les machines mobilisées par le jeu quand on fait référence à la démonstration du problème de l'arrêt dans sa présentation canonique telle qu'on l'a exposée plus haut [189].

La machine interne au jeu est en effet comparable à la machine E de la démonstration standard du problème de l'arrêt : ici la machine E duplique les instructions que l'on suppose être celle de l'homme et présente ses réponses en sortie à l'interrogateur. Cette machine E peut être imitée par une machine-*a* et donc aussi par une machine-*a* universelle. Elle peut donc être décrite par un ordinateur digital, comme Turing le souligne dans la description du jeu de l'imitation. Pour réussir à remplacer l'homme, il faut donc être capable d'*imiter* ses réponses et ses bluffs. Il y a donc bien ici, comme pour le cas du concept d'imitation, une extension métaphorique du concept d'imitation qui s'applique non plus à l'imitation de la table d'instructions d'une machine de Turing comme dans la théorie de la calculabilité mais à l'imitation des réponses dactylographiées de l'homme, interprétées comme des résultats des instructions d'une table de machine de Turing.

Pour ce qui est de la machine externe au jeu, celle que devient le lecteur-arbitre, de quel type de machine s'agit-il ? Cette machine doit, au vu des questions et des réponses dactylographiées interprétées comme les résultats des instructions d'une table de machine de Turing, parvenir à une décision quant à l'échec de l'interrogateur. En fait, la machine qu'il est nécessaire de postuler est celle que l'on a déjà rencontré dans la démonstration standard du problème de l'arrêt, la machine E*, qui s'arrête si E ne s'arrête pas sur l'entrée (description de la machine quelconque T) et qui ne s'arrête pas si E s'arrête sur l'entrée (description de la machine quelconque T). En effet, le lecteur-arbitre doit prendre mécaniquement la décision inverse de celle de l'interrogateur.

Mais on sait que, finalement, cette machine E* était *impossible à construire* et que c'était pour cette raison que le problème de l'arrêt se résolvait

---

[189] Cf. Première partie, chapitre 2, § 231.

par la négative. Le lecteur-arbitre devrait donc entrer dans un état de perplexité infinie, semblable à celui de l'interrogateur qui ne parvient pas à la bonne décision quant à la question du sexes des joueurs. Pourtant, le lecteur doit bien parvenir à une décision s'il veut opérer cette transformation de lui-même en machine; sans quoi, le jeu de l'imitation n'atteint pas son but qui est de dépasser l'illusion qui a trait à la différence entre les performances des êtres humains et celles des ordinateurs digitaux touchant la notion d'intelligence. La machine extérieure correspond donc exactement du cas de figure de la machine-*o,* machine capable par des moyens inconnus de parvenir à une décision sans que cette prise de décision soit entièrement mécanique. La machine extérieure doit donc être référée à l'oracle.

On voit donc qu'il y a une deuxième limite à l'analogie formaliste entre "On Computable Numbers …" et "Computing Machinery and Intelligence" concernant la notion d'imitation. L'imitation mécanique n'est pas équivalente à l'imitation ludique si l'on en reste au cas des machines-*a* : il faut réintroduire le cas de figure de la machine-*o* pour poursuivre intégralement l'analogie. Ce faisant, l'analogie perd tout son intérêt puisqu'elle ne permet plus d'identifier l'esprit à un fonctionnement mécanique au sens strict du terme.

Du point de vue du concept de décision comme du concept d'imitation, l'analogie formaliste entre la démarche de Turing du point de vue de la constitution de la théorie des machines de Turing et du point de vue de la constitution du projet de l'intelligence artificielle promet donc plus qu'elle ne peut réaliser. L'interprétation formaliste *présuppose* donc la validité de la thèse de Turing sur le cas de la notion d'intelligence alors que cette présupposition suppose résolue la difficulté résidant dans la différence entre machine-*a* et machine-*o*.

Curieusement, l'interprétation formaliste est pourtant la plus communément reçue et elle s'est popularisée grâce à une expression qui sert à décrire le jeu de l'imitation : celui-ci est couramment appelé "test de Turing". Ce test est censé permettre de déceler la présence d'intelligence chez les êtres humains et les ordinateurs. Nous allons essayer de montrer que cette expression est abusive.

2. Le "test de Turing" n'est pas un test mécanique

L'expression de "test de Turing" pour décrire le jeu de l'imitation est presque universellement répandue dans la littérature cognitive[190]. Elle constitue même une entrée dans la nouvelle édition du dictionnaire anglais Collins. Habituellement, la transformation du jeu de l'imitation en "test de Turing" entraîne une telle métamorphose des règles du jeu que sa forme originelle en devient méconnaissable.

Il est frappant de constater que seuls les logiciens-mathématiciens, pourtant les mieux placés pour "formaliser" les énoncés, semblent avoir remarqué que le jeu de l'imitation ne pouvait pas servir tel quel à la constitution d'un test permettant de démontrer la présence d'intelligence dans des entités physiques autres que l'être humain[191]. Revenons à l'article de 1950.

Le terme de test pour désigner le jeu de l'imitation est employé seulement trois fois dans "Computing Machinery and Intelligence" et à une page d'intervalle[192], pour répondre à l'objection que Turing appelle "l'argument de la

---

[190] Sans tenter d'être exhaustif, on la trouve à partir des années 70 (mais pas avant) sous la plume de Boden, Dennett, French, Ganascia, Haugeland, Hofstadter, Leiber, Michie, Penrose, Pylyshyn et Rastier respectivement dans **M. Boden "**Introduction" dans [**M. Boden ed.,** *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990], p. 4; **R. M. French**, "Subcognition and the limits of the Turing test", Mind, 99 (1990), p. 53-65; **J. Haugeland**, *Artificial Intelligence, the very idea,* trad. franç. *L'esprit dans la machine, Fondements de l'intelligence artificielle*, ed. Odile Jacob, Paris, 1989, p. 11; **D. Hofstadter** et **D. Dennett**, *The Mind's I* trad. franç. *Vues de l'Esprit*, InterEditions, Paris, 1987, p. 103; **J.-G. Ganascia**, *L'âme-machine; les enjeux de l'intelligence artificielle*, p. 203; **J. Leiber**, *An invitation to Cognitive Science*, Basil Blackwell, Oxford, 1991, chap. 9 "The Imitation Game"; **D. Michie**, *On Machine Intelligence*, a Halsted Press Book, John Wiley and Sons, New York, 1974, p. 65; **R. Penrose**, *The Emperor's New Mind*, Oxford University Press, Oxford, 1989, "The Turing test", p. 7-8; **Z. Pylyshyn**, *Computation and Cognition, Toward a Foundation for Cognitive Science,* op. cit., p. 53; F. **Rastier**, *Sémantique et recherches cognitives*, PUF, Paris, 1991, p. 31. Le biographe de Turing, Andrew Hodges, ne mentionne pas l'expression dans son livre, sauf comme entrée dans l'index sous la rubrique "Turing" déclinée ensuite en entrées multiples dont l'une s'intitule : "écrit un article classique sur l'intelligence mécanique avec le "jeu de l'imitation" comme Test de Turing". Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* Unwin Paperbacks, London, 1983, p. 584.

[191] Cf. **W. Hao**, *Popular lectures in mathematical logic*, Van Nostrand Reinhold Company, N-Y., p. 31 et **G. Kreisel**, "Review of Gödel's 'Collected Works, Volume I'", Notre Dame Journal of Formal Logic, 29, 1, Winter 1988.

[192] **A. M. Turing**, "Computing Machinery and Intelligence", *Mind*, op. cit., p. 446-447. Notons que Turing est l'auteur d'un test permettant la vérification des mémoires dans les ordinateurs ainsi que du premier texte portant sur la vérification de programme, qu'il écrivit à Manchester. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 365-366.

conscience"[193]. La généralisation de l'expression de "test de Turing" pour désigner le jeu de l'imitation nous paraît fausser la perception que l'on doit se faire du sens du jeu. L'expression de "test de Turing" a, dans sa forme grammaticale même, des incidences formalistes. En effet, on la rencontre habituellement en anglais sous la forme de l'expression [*Turing Test*] et non pas de l'expression [*Turing's test*], l'omission du cas possessif visant à accréditer l'aspect formel du jeu, qui se présente dès lors sous la même forme que celle d'un théorème mathématique, l'usage en anglais étant de nommer le théorème par le nom de son inventeur sans personnaliser le théorème par la marque du cas possessif (on parle ainsi de "machines de Turing" [*Turing machines*] sans employer de cas possessif). La dénomination du jeu par le terme de "test" est donc concomitante de sa dénaturation. On la trouve dès 1974 sous la plume de D. Michie qui fut pourtant un très proche collaborateur de Turing. Voici comment il décrit le jeu [194]:

> **«Le problème consistant à tester une machine pour savoir si elle est intelligente fut discuté pour la première fois par Alan Turing, grand logicien britannique et pionnier de l'ordinateur, qui mourut dans les années 50. […]. Turing proposa le test suivant. La machine devait être placée d'un côté d'un écran et un examinateur humain de l'autre côté. La conversation entre l'homme et la machine était rendue possible par le moyen d'une téléimprimante. Si après une heure ou deux de conversation imprimée, la machine avait été capable de tromper l'examinateur en lui faisant croire qu'il avait eu une conversation avec un être humain, alors, selon Turing, on devait concéder à la machine la prétention d'être intelligente».**

On reconnaît à peine, dans cette présentation du "test", le jeu de l'imitation tel qu'il a été formulé par Turing[195]. Remarquons que le terme de

---

[193] Cette objection lui avait été été adressée, comme il est dit dans l'article, par le professeur Jefferson. Le professeur Jefferson occupait la chaire de chirurgie neurale à l'Université de Manchester. Turing cite dans l'article la conférence de Jefferson (Lister Oration for 1949) sans en donner le titre : "L'esprit de l'homme mécanique" [*The Mind of the Mechanical Man*]. Turing participa également à un débat radiophonique contradictoire avec le même Jefferson, au studio de la BBC à Manchester, le 10 janvier 1952. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 404 et 450.

[194] Cf. **D. Michie**, *On Machine Intelligence*, op. cit., p. 66.

[195] Cependant, dans un article récent, le même D. Michie semble avoir été sensible à la nuance grammaticale touchant à l'usage du cas possessif; l'article, sans remettre en question la notion même de test, tente de montrer ses insuffisances et revient sur l'abus grammatical de l'expression de [*Turing test*]. Cf. **D. Michie**, "Turing's Test and conscious thought", Artificial Intelligence, 60 (1993).

"test" désigne habituellement, dans la littérature cognitive, la façon dont un organisme entre en contact actif avec son environnement par le biais d'essais et d'erreurs, par opposition à la façon passive dont il recueille des informations[196]. C'est ce sens technique de "test" qui semble avoir fait retour sur le texte de "Computing Machinery and Intelligence". L'article de 1950 laisse-t-il entendre que le jeu de l'imitation peut servir de test pour déceler la présence d'intelligence ?

Turing fait remarquer que la seule façon de s'assurer qu'une autre personne ou qu'une machine est bien susceptible de penser implique d'*être* la personne ou la machine en question[197]. Cette remarque impliquerait, dit Turing, une position sollipsiste que personne n'est prêt à adopter. Il faut donc, remarque Turing, accepter la convention "polie" qui attribue à autrui ou à une machine la possibilité de penser. On en conclut, bien que Turing ne le dise pas, qu'il n'est pas besoin de s'identifier à autrui ou à une machine pour lui supposer la capacité d'intelligence : les manifestations écrites de cette intelligence sont considérées comme des marques suffisantes de *comportement* intelligent. Le jeu de l'imitation permettant de vérifier la présence de ces marques, il peut de ce fait être considéré comme un test.

Mais on a vu que pour que le jeu de l'imitation puisse être considéré comme un test, il faut supposer l'existence d'un arbitre extérieur qui arrête le jeu. Sans cet arbitre extérieur, le jeu continue indéfiniment. Pour que le jeu soit considéré comme un test véritable, il faut donc rajouter une clause temporelle. Turing la mentionne comme en passant, quand il évoque l'avenir de l'intelligence artificielle et qu'il dit qu'une personne extérieure au jeu pourrait interroger l'interrogateur après cinq minutes de jeu [198]:

---

[196] Cf. **Z. Pylyshyn**, *Computation and Cognition, Toward a Foundation for Cognitive Science,* op. cit., p. 156-157.

[197] Turing dit refuser cette conception solipsiste. Comme nous l'avons montré cependant, c'est bien le but du jeu de l'imitation dans sa deuxième formule que de convaincre le lecteur que son mode de fonctionnement intellectuel n'est pas différent de celui d'une machine. Turing se plie donc, tout en la récusant, à cette conception solipsiste.

[198] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 442.

> **«Je crois que dans une cinquantaine d’années, il sera possible de programmer des ordinateurs […] de telle sorte qu’ils joueront si bien au jeu de l’imitation qu’un interrogateur moyen n’aura pas 70 chances sur cent de faire la bonne identification après cinq minutes de questionnement».**

Remarquons maintenant la chose suivante : pour parvenir à un arrêt, c’est-à-dire pour obtenir le même résultat que celui obtenu en faisant intervenir une personne extérieure au jeu tout en restant *à l’intérieur* du jeu, il faudrait s’accorder *l’infinité du temps,* puisque l’interrogateur n’est pas, pour Turing, censé réussir à faire la différence entre le sexe des joueurs. Or s’accorder l’infinité du temps revient à présupposer l’identification entre l’interrogateur et une machine de Turing possédant un ruban infini. Aussi peut-on voir dans la mention de l’attitude sollipsiste telle qu’elle est faite par Turing et sous réserve de se placer dans le jeu, non pas une attitude contraire au jeu de l’imitation mais au contraire la *règle* pour être certain de passer le “test de Turing”. Dans le cas où l’on se situe dans le jeu sans faire appel à l’infinité du temps, il n’y a donc pas de “test de Turing”, parce que cela présupposerait ce qu’il faut réussir à prouver : l’identification de l’être humain à une machine de Turing.

Maintenant, on peut objecter que Turing mentionne expressément que le jeu s’arrête au bout de cinq minutes par l’intervention d’un personnage extérieur au jeu. Mais où justifie-t-il la possibilité de cette intervention extérieure, c’est-à-dire d’un lieu extérieur au jeu ? *Nulle part*.

La question de la différence entre l’intérieur et l’extérieur du jeu est donc capitale. Une chose est dès à présent certaine : voir dans le jeu de l’imitation un “test” sans se préoccuper des conditions de sa réalisation est déjà une interprétation formaliste du jeu, qui présuppose sans la justifier l’identification du lecteur-arbitre à une machine de Turing. Pour réussir à se départir de cette interprétation formaliste, il faut donc réussir à comprendre comment Turing, tout en récusant l’attitude sollipsiste, ne justifie pas la différence qu’il opère entre l’intérieur et l’extérieur du jeu. C’est la réponse à cette question qui permettra finalement de comprendre si l’expression de “test” de Turing, appliqué au jeu de l’imitation, est ou non recevable.

3. L'origine du jeu de l'imitation

Turing a-t-il vraiment cherché à construire l'analogie formaliste sous la forme que l'on vient de décrire ? C'est ce qu'il faut essayer de savoir à présent.

Turing a mis plusieurs années avant de parvenir à créer le jeu de l'imitation tel qu'il apparaît en 1950. On trouve trace de son élaboration dans trois textes antérieurs : le premier est un rapport rédigé en mars 1946 pour le *National Physical Laboratory*, "Proposition pour le développement dans le département de mathématiques d'une machine automatique à calculer (ACE)"; le deuxième est une conférence prononcée à la *London Mathematical Society* qui date de février 1947 et le troisième, "Intelligent Machinery", est un rapport rédigé pour le *National Physical Laboratory* qui date de juillet et août 1948, c'est-à-dire au moment où Turing est nommé à l'université de Manchester. On va voir que Turing n'adopte pas clairement une perspective formaliste dans ces textes.

Lors de sa conférence de 1947, Turing décrivait comment devait s'organiser le travail avec un ordinateur, "l'ACE" [*Automatic Computing Engine*], qui était en cours de construction au *National Physical Laboratory* [199] :

> **«Grossièrement parlant, ceux qui travaillent en liaison avec l'ACE se divisent en deux catégories : ses maîtres et ses serviteurs. Ses maîtres imaginent des tables d'instruction pour lui, en réfléchissant de plus en plus profondément aux façons de l'utiliser. Ses serviteurs le nourrissent de cartes au moment où il les demande. Ceux-ci répareront toute partie qui deviendrait défectueuse. Ils assembleront les informations dont il a besoin. En fait, les serviteurs remplaceront les membres. Au fur et à mesure que le temps passera, le calculateur [*calculator*] lui-même prendra les fonctions des maîtres et des serviteurs. Les serviteurs seront remplacés par des membres mécaniques et électriques et des organes des sens. On pourra par exemple fournir des suiveurs de courbes pour faire en sorte que les informations soient directement prélevées sur les courbes plutôt que d'avoir des filles pour lire les valeurs et les reporter sur des cartes. Les maîtres seront susceptibles d'êtres remplacés, parce que dès qu'une technique quelconque deviendra un tant soit peu stéréotypée, il deviendra possible de constituer un système de tables d'instructions qui rendra le calculateur [*computer*] électronique capable de l'exécuter pour lui-même. Il pourrait se produire cependant que les maîtres refusent de le faire. Ils renâcleront à l'idée de se faire voler leur travail de cette façon. Dans ce cas, ils entoureront tout leur travail de mystère […]».**

Un détail doit être éclairci pour que le texte devienne compréhensible : quelles sont donc ces "filles" dont Turing ne précise pas l'existence à aucun moment de sa conférence ?

---

[199] **A. M. Turing**, "Lecture to the London Mathematical Society on 20 February 1947", reprint dans **B. E. Carpenter et R.W. Doran eds.,** op. cit., p. 121

On peut supposer qu'il s'agit d'un souvenir de l'époque immédiatement antérieure à la construction de l'ACE. Pendant toute la durée de la guerre, Turing a travaillé au centre de décryptage des messages codés de l'armée allemande : le travail consistait, à l'aide d'un certain nombre de machines utilisant des cartes perforées, à déchiffrer les messages radio des sous-marins allemands et à prévoir leur position pour que les convois de bateaux britanniques chargés d'acheminer en Angleterre les matériaux nécessaires à l'effort de guerre ne soient pas coulés en mer. Or l'équipe des cryptographes était essentiellement masculine[200] tandis que celle qui confectionnait les cartes perforées était exclusivement composée de femmes appartenant au WRNS[201]. Les équipes de cryptographes les appelaient habituellement les “filles”[202]. On comprend alors l'allusion étrange de Turing à des “filles” chargées de transcrire les valeurs représentants des points des courbes sur des cartes perforées : les programmes de l'ACE étaient précisément rédigées sur ce type de carte.

L'intérêt du texte de la conférence de 1947 vient de ce que l'on y reconnaît les deux étapes de la constitution de la notion d'intelligence en concept universel, constitution qui passe par l'abolition de la différence des sexes. L'avenir du travail de la machine est décrit en effet comme consistant à abolir la différence des sexes. Cette abolition exige deux étapes.

Dans la première, les femmes servent les hommes qui servent l'ordinateur. Quelle est la nature du travail des femmes ? Quand on se reporte à un texte tout juste antérieur à la conférence de 1947, le “Proposition pour le développement dans le département de mathématiques d'une machine automatique à calculer (ACE)” de 1946, on se rend compte qu'à cette époque, Turing considérait que le problème de l'intégration de l'aire sous une courbe était insoluble, parce qu'on ne

---

200 Nous reviendrons plus loin sur la présence de femmes dans l'équipe des cryptographes.

201 WRNS : Women's Royal Naval Service. Service Féminin de la Marine Royale.

202[*the girls*]. Cf. **I. J. Good**, “Introductory Remarks for the article in Biometrika 66 (1979) “A. M. Turing's Statistical Work in World War II” publié dans [**A. M. Turing**, *Collected Works of A. M. Turing*, vol. 1 “Pure Mathematics”, North-Holland, 1992], p. 215-217. Hodges précise qu'elles étaient appelées [*big room girls*]. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 195.

pouvait fournir à la machine les données appropriées[203] alors que, lors de sa conférence de 1947, Turing suggérait, pour le même problème, de remplacer les “filles”, dont il *omettait* de préciser le rôle mais qui était sûrement celui qu’elles avaient pendant la guerre, par des mécanismes automatiques permettant de donner une approximation digitale des courbes. Turing accorde tout d’abord aux femmes un rôle de maintenance corporelle : elles sont à la fois celles qui «nourrissent» la machine et celles qui en remplacent les «membres». Le rôle des femmes est donc de fournir à l’ordinateur une existence matérielle. Elles sont seulement des intermédiaires *analogiques* de ce que l’ordinateur sera capable dans l’avenir de traiter de façon *digitale* grâce au remplacement des organes physiques des servantes par des appareils mécaniques ou électriques. Les femmes sont donc rangées du côté de la matière physique, par opposition à l’homme associé au travail intellectuel d’une part et à l’ordinateur décrit comme machine digitale d’autre part. La matière physique continue associée aux femmes devient matière physique discrète associée à l’ordinateur par le biais du travail d’abolition de la différence des sexes mis en place par les hommes.

Dans une seconde étape décrite au futur, correspondant au jeu de l’imitation n°2, la différence des sexes (et la division du travail qui lui est liée) sera elle-même abolie, puisque le calculateur sera capable de remplacer le travail de l’homme. Notons que le terme utilisé pour décrire la fonction du calculateur lors de cette seconde phase est celui de [*Calculator*] et non celui de [*Computer*]. Le “calculateur” n’est référé ni à la *machine* qu’est l’ACE, ni à l’être humain : le statut, mécanique ou humain, du “calculateur” est ainsi laissé dans l’ombre, comme le requiert le second jeu de l’imitation.

Le texte de la conférence de 1947 s’achève sur une liste d’objections, comme c’est le cas dans “Computing Machinery and Intelligence”. L’objection la plus forte pour Turing est qu’il y aurait une contradiction fondamentale entre l’idée d’intelligence et celle de machine. Turing se réfère à son travail de 1936

---

[203] Cf. **A. M. Turing**, “Proposal for Development in the Mathematics Division of an Automatic Computing Engine (ACE)”, op. cit., section 8, problème 7, reprint dans **B. E. Carpenter et R.W. Doran eds.,** op. cit., p. 41 : «Il ne serait pas possible d’intégrer l’aire sous une courbe, car la machine n’aurait pas l’entrée [input] appropriée».

pour la réfuter :

> **«On a par exemple montré que dans le cas de certains systèmes logiques, il ne peut pas y avoir de machine qui pourrait distinguer les formules démontrables du système des formules indémontrables; c'est-à-dire *qu'il ne peut pas y avoir de test* [204]qu'une machine appliquerait et qui pourrait diviser avec certitude les propositions entre ces deux classes. Si une machine est construite pour ce but, elle doit sur certains cas ne pas pouvoir donner de réponse.**

Dans son analogie avec la perspective de "On Computable Numbers …", Turing précise que, de ce point de vue déjà, *il ne pouvait y avoir de test* pour diviser les formules d'un système formel en deux classes, démontrables et non-démontrables. On pourrait ajouter : si c'est le cas du point de vue formel, c'est aussi le cas dans une perspective non-formelle comme celle qui a trait à la caractérisation de la notion d'intelligence. L'idée de "test de Turing" est donc un contresens du point de vue de la perspective formaliste elle-même.

Turing précise dans la suite de la conférence qu'au lieu de faire en sorte que, sur certaines entrées, la machine ne donne pas de réponse, on peut faire en sorte qu'elle donne par moments des réponses fausses, comme le ferait tout mathématicien humain. Ce mathématicien est décrit au masculin :

> **«Mais le mathématicien humain, de la même manière, ferait des erreurs en essayant de nouvelles techniques. Il nous est facile de considérer ces erreurs comme sans importance et de lui [*him*] donner une autre chance, mais on accordait sans doute aucune pitié à la machine. En d'autres termes donc, on s'attend à ce que si une machine soit infaillible, elle ne puisse pas aussi être intelligente».**

Ici encore, le jeu de l'imitation mettra ouvertement en place ce que la conférence de 1947 ne fait que décrire inconsciemment : le mathématicien est un homme - et non pas un être humain - qui finira par être éliminé si l'on accorde à la machine la même chose qu'au mathématicien : le droit à l'erreur. C'est dans l'interaction et l'entraînement, comme le pose la règle du jeu de l'imitation, que ces erreurs peuvent être progressivement éliminées et que l'on peut à bon droit parler de "test". Immédiatement après la dernière citation que nous venons de donner dans laquelle Turing montre que la faillibilité n'implique pas l'absence

[204] C'est moi qui souligne.

d'intelligence, celui-ci remarque :

> **«Pour continuer ma plaidoirie en faveur d'une attitude de fair-play à l'égard des machines quand on teste leur Q. I. : un mathématicien humain a toujours eu à subir un entraînement intensif. Cet entraînement peut être considéré comme n'étant pas différent de l'introduction des instructions dans une machine. […] Personne ne contribue beaucoup à l'avancement des connaissances, pourquoi en attendre plus de la part des machines ? Autrement dit, on doit accorder à la machine la possibilité d'avoir des contacts avec des êtres humains de sorte qu'elle puisse s'adapter à leur niveau. Le jeu d'échecs peut peut-être servir à ce but, dans la mesure où l'adversaire de la machine établit automatiquement ce contact».**

Le test dont parle Turing ici n'est donc pas un test mécanisable et il ne dérive donc pas de la perspective mise en place dans "On Computable Numbers …". Seul le modèle du jeu peut en rendre compte parce que, contrairement à ce qui pourrait apparaître à première vue, le jeu n'est pas équivalent à un système formel, dans la mesure précisément où il ne suppose pas constitué le terrain du formel lui-même mais une situation d'interaction. C'est pourquoi le modèle du jeu ne doit pas être envisagé dans une perspective formalisante, même dans le cas où il s'agit d'un jeu très "formel" comme le jeu d'échecs.

Dans un texte ultérieur, Turing revient sur l'exemple d'une partie d'échecs dans laquelle l'un des adversaires serait un ordinateur. La règle du jeu qu'il décrit ne porte pas sur la détermination des sexes des joueurs puisque ce sont tous des hommes [205]:

> **«Il n'est pas difficile de construire une machine de papier qui pourra jouer une partie d'échecs pas trop mauvaise. Prenons trois hommes [*men*] pour l'expérience, A, B, C. A et C doivent être plutôt de mauvais joueurs, B est l'opérateur qui fait fonctionner la machine de papier. (Pour qu'il puisse la faire fonctionner de façon relativement rapide, il est recommandé qu'il soit à la fois mathématicien et joueur d'échecs). Deux pièces sont utilisées ayant certaines dispositions permettant la communication des coups et C joue une partie soit contre A soit contre la machine de papier. C peut trouver difficile de dire contre qui il joue. (Ceci est une forme assez idéalisée d'une expérience que j'ai véritablement menée).»**

C joue donc aux échecs soit contre un homme soit contre une créature humaine mais servile, la "machine de papier". C peut-il réussir à déterminer si son adversaire est un être humain "mécanisé" ou un être humain tout court ? On reconnaît ici le canevas du jeu de l'imitation et on peut à bon droit se demander ce

---

205 **A. M. Turing**, "Intelligent Machinery", op. cit. p. 51.

qui a poussé Turing, dans les deux ans qui ont suivi, à modifier la règle du jeu en abandonnant le paradigme du jeu d'échecs (qui n'est plus qu'un exemple parmi d'autres) et à donner au jeu la forme qu'on lui connaît dans "Computing Machinery and Intelligence".

En fait, en abandonnant l'exemple de la partie d'échecs tout en gardant le montage expérimental qui entoure le jeu, Turing, dans "Computing Machinery and Intelligence", revient à la perspective qui était la sienne en 1947. Si Turing a modifié le jeu, c'est qu'il y a trouvé un avantage eu égard au but qu'il cherche à atteindre, à savoir constituer la notion d'intelligence en un concept universel qui serait applicable à d'autres cas qu'au cas humain. De ce point de vue, le jeu de l'imitation constitue un réel progrès, dans la mesure où c'est bien la différentiation à opérer entre l'intellectuel et le physique qui occupe le premier plan de l'analyse. Au contraire, dans le cas du jeu d'échecs de "Intelligence Machinery", on ne distingue pas clairement le physique de l'intellectuel, puisque l'adversaire humain et l'adversaire mécanique ne sont pas susceptibles d'être distingués du point de vue de leur apparence physique : rien ne distingue extérieurement le joueur humain du joueur humain *servile*. En revanche, si l'on plaque la différence des sexes sur cette la différence entre joueur humain servile et joueur humain non-servile, l'aspect physique du problème ressort en pleine lumière.

Le jeu de l'imitation apparaît donc comme la synthèse des trois textes antérieurs. Il doit permettre de résoudre cet apparent paradoxe : il faut d'une part parvenir à distinguer *physiquement* les joueurs et d'autre part montrer que cette différence n'a *aucune incidence* sur le déroulement du jeu. Ce sont précisément les conditions que le jeu de l'imitation tente de réunir, mais la rançon à payer est l'établissement d'une différence entre l'intérieur et l'extérieur du jeu, différence inassignable en termes de machine-*a* puisqu'il faut aussi invoquer la notion de machine-*o* pour décrire la situation du jeu et le but ultime recherché, à savoir la conversion "en esprit" du lecteur en machine.

4. Caractère intellectualiste de l'interprétation formaliste

La différence à effectuer entre le formel et le ludique a trois conséquences qui s'enchaînent. Premièrement, l'analogie entre le cas de la calculabilité et le cas du jeu de l'imitation est inadéquate. Deuxièmement, la notion d'intelligence n'est pas un *concept* abstrait, concept qui serait obtenu une fois qu'il aurait été prouvé par le jeu de l'imitation que les caractéristiques physiques des joueurs sont sans importance. Troisièmement, le fondement de la psychologie scientifique, que l'on tentait de baser sur une application au cas de l'intelligence des résultats mathématiques touchant le concept de calculabilité, devient problématique. C'est cette dernière conséquence que nous allons étudier maintenant.

41. Le mirage de l'auto-fondation de la thèse de Turing

Pour l'interprétation formaliste, puisque la thèse Turing, confondue avec la thèse de Church, permet de définir formellement l'universalité de la notion d'algorithme par le biais du concept de machine, il est peut-être possible d'étendre la thèse en question à d'autres notions, générales mais imprécises, comme celle d'intelligence. Mais le choix de la notion d'intelligence, qui fait l'objet de l'investigation menée dans “Computing Machinery and Intelligence”, n'est pas fortuit.

La notion d'intelligence n'est en effet pas une notion comme les autres, dans la mesure où elle entretient avec la notion de psychologie des liens étroits. Aussi, en tentant de préciser la notion d'intelligence par le biais d'une généralisation de la thèse de Turing, le sens de la thèse s'est-il modifié : la thèse ne vise plus seulement à assurer le passage entre une notion intuitive et une notion mathématique par le biais d'une analyse psychologique; elle vise, par une espèce de retour sur soi, à se fonder formellement elle-même *en rendant précise la notion même de psychologie*. Mais un tel but est impossible à atteindre : Turing, usant *déjà* d'une argumentation de nature psychologique pour justifier une thèse mathématique (la thèse de Turing), il est exclu que le domaine du psychologique puisse, en dernière instance, trouver son fondement dans cette thèse mathématique. Il y aurait là un cercle vicieux.

Une conséquence directe s'en suit : il y a bien une *preuve formelle* du caractère légitime de l'imitation mathématique entre les algorithmes mais il *ne peut pas y avoir* de preuve formelle du fait que l'imitation “ludique” permette d'identifier l'intelligence humaine et l'intelligence mécanique, puisque cette imitation, de nature psychologique, ne peut pas faire appel à une thèse psychologique comme celle de Turing pour être fondée. Dès lors, l'imitation dans le jeu ne pourra jamais faire l'objet d'une définition formelle, contrairement à l'imitation prise au sens mathématique dans la théorie des algorithmes.

42. Le piège linguistique que représente le terme d'intelligence

Il y a donc, quand on parle d'intelligence, un piège linguistique dans lequel il ne faut pas tomber : en prononçant le terme d'intelligence", on est induit à penser à un objet *intellectuel*, alors qu'il s'agit là d'une pétition de principe. C'est pourquoi Turing considère que la question "les machines peuvent-elles penser ?" ne mérite pas qu'on s'attarde sur elle [206]: en fait, cette pseudo-question philosophique suppose déjà le problème résolu[207]. Le problème véritablement philosophique que Turing tente de résoudre est plus fondamental, parce qu'antérieur à toute spéculation sur la question de la définition de l'intelligence conçue indépendamment de tout substrat physique : il s'agit d'étudier les conditions de possibilité de *toute* réflexion sur le concept d'intelligence, conditions qui passent, pour Turing, par la distinction à opérer entre le domaine du physique et celui de l'intellectuel. L'universalité qui est visée dans "Computing Machinery and Intelligence" n'est donc pas de même nature que dans "On Computable Numbers", qui se situait déjà sur le terrain intellectuel et sur lui seulement. Il s'agit, dans le cas de l'article de 1950, d'une universalité qui porte sur le *rapport* entre l'intellectuel et le physique, universalité proprement philosophique. C'est donc sur la distinction entre le physique et l'intellectuel que Turing travaille, pour tenter, *par la suite*, mais par la suite seulement, de définir le concept général d'intelligence, hors de toute préoccupation concernant le substrat physique. Et ceci est bien, en effet, le but ultime de Turing : s'il y a une *objectivité* de l'intelligence, c'est que celle-ci ne dépend pas des caractéristiques particulières propres à l'intelligence humaine, qui ne sont que le produit particulier de l'évolution de l'espèce au cours du temps.

Si le jeu de l'imitation ne vise pas à *prouver* de manière irréfutable l'intelligence de la machine, arrive-t-il cependant à *montrer* qu'il est possible de distinguer le physique de l'intellectuel ? Il faut, pour résoudre la question qui a été, jusqu'à présent, seulement posée, réussir à déterminer quels rôles spécifiques

---

206 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 442.

207 Cette question "rhétorique" fait cependant l'objet du titre de la conférence radiodiffusée que Turing donna à la BBC le 15 mai 1951. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 441.

jouent l'homme et la femme dans leur rapport à l'ordinateur au sein des deux formules du jeu de l'imitation.

Chapitre III

Interprétation probabiliste du jeu de l'imitation

Il y a un aspect important de l'argumentation de Turing que l'interprétation formelle du jeu de l'imitation ne prend pas en compte, c'est son aspect probabiliste. Turing fait en effet intervenir un argument de ce type pour essayer d'emporter la conviction du lecteur quant à l'issue finale du jeu. L'usage de probabilités permet-il de répondre à la question fondamentale à laquelle nous nous étions déjà heurtés concernant la possibilité pour l'interrogateur de distinguer, à partir du point de vue interne au jeu, le physique de l'intellectuel ? L'intérêt de l'argument probabiliste vient du fait qu'*il impose une échéance temporelle* pour répondre à la question alors que ce n'était pas le cas de l'argument formaliste.

Après avoir exposé la nature de l'argument utilisé par Turing, nous essayerons de nous demander si celui-ci est suffisant pour atteindre le but qu'il a fixé au jeu de l'imitation.

1. Aspect probabiliste de l'issue du jeu de l'imitation

Turing, dans la section 6 de “Computing Machinery and Intelligence”, accorde un grand rôle au temps dans la justification de l'argumentation proposée par le jeu de l'imitation : le temps passant, le public cultivé finira par s'habituer à l'idée que l'intelligence est un concept universel qui peut aussi s'incarner dans des ordinateurs. Après Turing, on a usé et abusé de cet argument pour vanter les mérites de l'intelligence artificielle en imaginant que, dans l'avenir, des ordinateurs seront capables de faire “mieux” que les êtres humains dans des tâches de moins en moins mécaniques[208], ce qui suppose que l'on possède un critère intemporel permettant de distinguer le "mieux" du "moins bien". Habituellement, ce critère est entièrement passé sous silence, contrairement à ce qui se produit dans le jeu de l'imitation.

L'argument de Turing ne porte pas sur les futures réalisations de l'intelligence artificielle bien qu'il prenne en compte la possibilité d'un progrès technologique. Le point de vue de Turing, fidèle en cela à la règle du jeu de l'imitation, porte exclusivement sur le changement d'attitude du public tel qu'il pourra se manifester dans le langage.

L'argument probabiliste est le suivant [209]:

> **«Je crois que dans une cinquantaine d'années, il sera possible de programmer des ordinateurs […] de telle sorte qu'ils joueront si bien au jeu de l'imitation qu'un interrogateur moyen n'aura pas 70 chances sur cent de faire la bonne identification après cinq minutes de questionnement. […] Je crois qu'à la fin du siècle, l'usage des mots et l'opinion générale des personnes éduquées aura changé si complètement que l'on sera capable de parler de machines qui pensent sans s'attendre à être contredit».**

Turing définit donc dans son argument *deux* limites temporelles.

La première est interne au jeu : c'est celle qui a trait à la durée d'une partie, fixée en l'occurrence à cinq minutes. On peut concevoir que la durée d'une partie serait fixée avant de commencer et qu'il s'agirait donc là d'une nouvelle règle, connue des participants au jeu.

La seconde est externe : c'est celle qui a trait à la validation de l'expérience du jeu et qui est fixée à un siècle. Turing suppose de plus trois étapes échelonnées sur le siècle pour que son argumentation finisse par être acceptée : état de la question en 1950 - date à laquelle il rédige l'article -, puis état de la

---

[208] Un point de vue caricatural qui confine à l'escroquerie pure et simple (mais qui est symptomatique) est celui de Jastrow dans **R. Jastrow**, *Au-delà du cerveau*, traduction française, Hachette, 1984, p. 209 : «Depuis le début des années 1950 et la naissance de l'ordinateur moderne, les machines ont connu une formidable augmentation de leur puissance et de leurs capacités. La première génération d'ordinateurs était un milliard de fois plus encombrante et moins efficace que le cerveau humain. Aujourd'hui, l'écart qui les sépare est divisé par mille. Et il sera peut-être entièrement comblé aux alentours de 1995. La courbe ascendante de l'intelligence artificielle ne connaît pas de limites; contrairement au cerveau humain, les ordinateurs n'ont pas à subir l'épreuve de la naissance». La pseudo-quantification des "capacités" humaines et de celles des ordinateurs donne un vague tour scientifique à ces propos inspirés. Les dernières pages du livre passent d'ailleurs de la métaphysique scientiste au récit de science-fiction.

[209] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 442.

question en l'an 2000, enfin état de la question en 2050. Au fur et à mesure de l'écoulement du temps, la probabilité pour qu'un interrogateur moyen ne parvienne pas à opérer les bonnes identifications pour une durée donnée du questionnement augmente. En 2050, il sera admis que l'échec de l'interrogateur est total et qu'il n'a aucune chance de parvenir à opérer les bonnes identifications. Bien que ce ne soit pas mentionné expressément, Turing a sans doute à l'esprit un schéma du type : en 1950, l'interrogateur a 50% de chances d'échec, 70% en l'an 2000, 100% en 2050. Examinons ces étapes.

Turing ne mentionne pas les résultats de la première étape, celle de 1950, mais le nombre de 50% correspond en fait à des réponses données au hasard : en effet, en répondant au hasard, l'interrogateur a une chance sur deux de se tromper dans ses identifications, que ce soit dans le jeu n°1 ou dans le jeu n°2.

Lors de la seconde étape, en l'an 2000, il y aura encore besoin, selon l'argument de Turing, de montrer expérimentalement, par le jeu de l'imitation et au moyen des ordinateurs que l'on concevra à cette époque, que le concept d'intelligence a relativement peu de chances d'être lié au substrat physique particulier de l'espèce humaine. Mais on sera déjà passé d'une situation aléatoire à une situation ordonnée.

A la fin de la troisième étape, en 2050, toute résistance de la part du public éduqué sera censée avoir disparu et il sera entré dans les mœurs d'accorder aux ordinateurs digitaux ce qui semblait être encore en 1950 l'apanage des seuls êtres humains : leur intelligence. La prise en compte d'une évolution temporelle semble donc importante au bon déroulement de la preuve non-formelle qu'est censée fournir le jeu de l'imitation. Comment concevoir cette perspective temporelle ?

La différence entre le point de vue interne et le point de vue externe se manifeste de façon temporelle. Ainsi cette différence est-elle reconnue et non plus présupposée comme c'était le cas dans le pseudo-test de Turing. Il y a de plus une correspondance entre les deux points de vue temporels : pour une durée de jeu inchangée (cinq minutes), l'interrogateur voit ses chances de réussite diminuer et le lecteur-arbitre voit les chances d'identifier être humain et ordinateur augmenter. Les deux fonctions convergent vers un point qui doit, selon l'argument, être atteint en 2050, point où la différence entre le point de vue intérieur et le point de vue extérieur est pleinement établie.

Remarquons que cette convergence est postulée : Turing veut donc convaincre le lecteur qu'il est *dès maintenant* possible de se placer *en esprit* à la limite temporelle de 2050, moment où l'on pourra concevoir qu'il n'y a pas *en réalité* de différence entre les êtres humains et les ordinateurs digitaux. Le déplacement opéré en esprit jusqu'en 2050 a donc pour conséquence *d'abolir la perspective temporelle* et de faire parvenir à une conclusion objective qui ne dépend pas du temps, à savoir que l'intelligence est un concept universel qui peut s'incarner dans des substrats divers. Le déroulement du temps joue donc ici comme un facteur de désillusion progressive quant à la pseudo-différence entre les êtres humains et les ordinateurs digitaux d'une part et favorise la prise de conscience de la différence réelle entre le physique et l'intellectuel d'autre part. Autrement dit, du point de vue externe au jeu, l'argument vise à convaincre le lecteur que l'absence de différence entre être humain et ordinateur digital, encore concevable en esprit au début du jeu, doit disparaître à la fin en réalité. En quoi le fait de se projeter en esprit jusqu'en 2050 peut-il réussir à opérer cette conversion ? Il faut, pour tenter de répondre à cette question, décrire la façon dont Turing envisageait la mécanisation des jeux et sa participation à la construction des premières machines à jouer.

2. Turing et la mécanisation des jeux

Turing s'est intéressé à la mécanisation des jeux parce qu'il rapprochait cette activité de celle de mécanisation de la pensée. Cette analogie s'est construite progressivement : alors qu'il compare la pensée du calculateur au déroulement d'un algorithme dans “On Computable Numbers …”, il associe après la guerre l'activité de pensée au déroulement d'un jeu. On va voir que le modèle du jeu introduit dans l'analogie avec la pensée une différence capitale parce que la notion de jeu constitue une généralisation et une complexification par rapport à la notion d'algorithme.

La distinction première à partir de laquelle on peut construire une typologie des jeux est celle qui distingue les jeux à une personne ou puzzles des jeux à plusieurs personnes. En effet, trouver la solution d'un puzzle revient à appliquer les étapes d'un algorithme : chaque étape est entièrement déterminée par la précédente et ce, jusqu'à la solution. Ce n'est pas le cas d'un jeu à plusieurs personnes dans lequel plusieurs options s'offrent à chaque étape, ce qui nécessite l'élaboration d'une stratégie définie en termes de probabilités de gain[210]. On voit donc que le cas du jeu à plusieurs est une généralisation du cas du puzzle, généralisation qui exige l'introduction d'un calcul de probabilités. En conséquence, le modèle du jeu à plusieurs introduit aussi une généralisation dans l'analogie avec le fonctionnement de la pensée. L'étude de ces distinctions doit permettre de mieux comprendre la place du jeu de l'imitation et le rôle qu'il joue dans la caractérisation de la notion de psychologie.

21. Jeux à une personne : les puzzles

Dans un article "grand public", "Solvable and Unsolvable Problems"[211], postérieur à "Computing Machinery and Intelligence" mais qui résume les acquis réalisés depuis "On Computable Numbers …" touchant les questions de décision, Turing décrit un certain nombres de puzzles et montre que la résolution de tout puzzle consiste à appliquer un algorithme de résolution au problème posé. Turing commence par dire qu'il n'y a pas de test qui permettrait de savoir si tout puzzle est soluble ou pas, reprenant ainsi l'acquis majeur de son article de 1936[212]. Il montre ensuite que la notion de puzzle ne doit pas être considérée comme un

---

[210] Comme le fait remarquer D. Michie dans **D. Michie**, *On Machine Intelligence*, a Halsted Press Book, John Wiley and Sons, New York, 1974, p. 21 : «Le caractère distinctif du puzzle, en tant qu'il s'oppose au jeu, est que le changement effectué […] par un coup donné est entièrement déterminé. Dans un jeu en revanche, le joueur doit prendre en considération non pas une conséquence nécessaire unique de son coup mais un éventail de conséquences possibles alternatives, dont une sélection sera opérée par le coup ou les coups de son adversaire avant que ce ne soit à nouveau à son tour de jouer».

[211] **A. M. Turing**, (1954), "Solvable and Unsolvable Problems", Science News, 31, Penguin Science, pp. 1- 23.

[212] Il propose en particulier une nouvelle démonstration du problème de l'arrêt adaptée au cas des puzzles.

simple amusement parce qu'elle peut servir de représentation à celle d'algorithme, si importante en logique mathématique :

«La tâche qui consiste à démontrer un théorème mathématique dans le cadre d'un système axiomatique est un très bon exemple de puzzle».

Ainsi l'obtention d'un théorème dans un système formel est-il réductible à la solution d'un puzzle. Il devient alors possible d'énoncer une variante de la thèse de Turing pour les puzzles :

«Pour tout puzzle donné, nous pouvons trouver un puzzle de substitution qui lui corresponde et qui lui est équivalent au sens où, si une solution est donnée pour celui-ci, nous pouvons facilement trouver une solution pour l'autre».

Ce puzzle de substitution correspond en fait à la mise sous forme d'algorithme susceptible d'être traité par une machine de Turing. Ainsi la solution de tout puzzle peut-elle être effectuée par une machine de Turing, si une solution pour le puzzle existe.

On pourrait être tenté de faire une analogie entre ce type de démarche et celle du jeu de l'imitation : le dédoublement entre le jeu n°1 et le jeu n°2 proviendrait de la solution que Turing a donnée au problème de la décision dans le cadre des systèmes formels. Le jeu n°2 jouerait le rôle de puzzle de substitution, susceptible de recevoir une formulation en termes de machines. *Il y aurait alors moyen de transformer le jeu de l'imitation en puzzle*. On peut alors présenter sous cette forme le problème de l'interprétation de la viabilité du jeu de l'imitation : la réduction de ce jeu à un puzzle est-elle ou non possible ? La réponse à cette question permettrait de préciser le modèle de la psychologie qui peut découler du jeu. En particulier, c'est le statut de la notion de représentation qui est en question ici : la notion de représentation est inutile dans le cas du puzzle[213] alors qu'elle est indispensable dans le cas du jeu dans la mesure où chaque joueur doit se faire une

---

[213] On peut, à partir du moment où l'on représente le déroulement d'un algorithme comme le déroulement des étapes de la solution d'un puzzle, se demander s'il est nécessaire d'invoquer la “pensée” du calculateur pour rendre compte de la mécanisation de la notion d'algorithme : celle de puzzle et de puzzle de substitution, moins psychologique, suffirait. Cf. **W. Hao**, “Games, Logic and Computers”, <u>Scientific American</u>, 1965, 213 (5), pp. 98-106, dans lequel l'auteur décrit la mécanisation des algorithmes par le biais de jeux de dominos de différentes couleurs pavant le plan.

représentation virtuelle des conséquences de son ou de ses coups futurs. Cette représentation virtuelle s'appelle une *stratégie*. Examinons cette notion et sa mécanisation possible.

22. Jeu à plusieurs personnes

Turing a travaillé sur la notion de jeu en favorisant trois axes de recherche[214] : l'évaluation des gains en termes de minimax, la notion de "points morts" et l'apprentissage des machines. Étudions ces trois notions avant de voir comment Turing en a tiré profit pour l'élaboration du jeu de l'imitation.

221. Évaluation d'un jeu en termes de minimax

On peut représenter un jeu sous forme d'un arbre de décision et attribuer des poids d'équivalence à chaque ramification représentant une décision stratégique. Si l'on fait l'hypothèse que chaque joueur suit une stratégie optimale, on peut prouver que l'on peut assigner des valeurs numériques non seulement aux points terminaux de l'arbre mais aussi, en "remontant" les branches de l'arbre, à tous les points de décision, y compris le premier. On a montré qu'il était possible de faire effectuer à une machine cette recherche d'une évaluation en termes de minimax. Comme le remarque D. Michie :

> **«Si l'on a les moyens d'attacher des valeurs numériques à n'importe quel point, nous avons un schéma simple pour une machine capable de jouer à des jeux : regarder l'arbre jusqu'à une certaine profondeur, évaluer tous les points à cette profondeur, puis faire "remonter" l'évaluation par la méthode du minimax […]».**

La difficulté de l'évaluation en minimax est celle de l'explosion combinatoire : pour certains jeux, comme les échecs par exemple, le nombre de

---

[214] C'est ce que rapporte D. Michie dans **D. Michie**, *On Machine Intelligence*, op. cit., p. 36 sq. Celui-ci cite l'article rédigé par Turing, "Digital Computers Applied to Games", dans le livre de référence concernant les premiers développements des ordinateurs britanniques, *Faster Than Thought*, édité par Bowden. C'est en 1949, soit un an avant "Computing Machinery and Intelligence", rapporte A. Hodges, que Turing a abandonné, faute d'une reconnaissance minimale de la valeur de son travail par le même Bowden, le désir de se tenir au courant des développements technologiques les plus récents touchant les ordinateurs. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 478.

parties virtuelles est énorme et il est impossible de chercher à parcourir la totalité de l'arbre.

Turing a montré comment on pouvait faire intervenir dans l'évaluation un certain nombre de données stratégiques générales qui évitent d'avoir à retracer l'intégralité des points de décision avant d'adopter une stratégie. Dans le jeu d'échecs par exemple, on peut faire intervenir dans l'évaluation des critères comme la mobilité des pièces, le contrôle des cases centrales, l'avancement des pions et tout autre aspect qui peut sembler important compte tenu du jeu choisi. Cette "stratégie compressée", selon l'expression de D. Michie, est importante dans la mesure où elle offre un analogon de la procédure psychologique de généralisation et d'abstraction tout en étant susceptible d'être mécanisée : le choix de ces critères de jeu peut être intégré à un programme et peut donc servir à la construction d'une stratégie par une machine. Remarquons cependant que la machine en question ne construit pas ces données stratégiques générales mais qu'elle les reçoit et se contente de les appliquer, quelque fois beaucoup mieux que les êtres humains qui en sont pourtant les inventeurs[215].

Un autre moyen d'éviter l'explosion combinatoire est de restreindre la représentation virtuelle des coups à venir et de s'en tenir à un niveau de profondeur virtuelle limité. C'est à la définition de cette profondeur que Turing a réfléchi en inventant la notion de "point mort".

222. La notion de "point mort"

Il s'agit d'une notion qui permet de limiter l'exploration de l'arbre de décision à ses branches les plus aptes à donner des résultats. Par exemple aux échecs, un "point mort" est un point de l'arbre à partir duquel les occasions de capture des pièces de l'adversaire sont nulles, au moins pour la profondeur de la représentation virtuelle des coups à laquelle on s'est fixé. La notion de point mort fait donc intervenir l'idée de profondeur variable dans la représentation virtuelle.

[215] D. Michie rapporte le cas du premier programme effectif de jeu d'échecs, le "Samuel Checker Program" qui battait en 1959 son propre inventeur, A. L. Samuel, dans une proportion supérieure à 50%. Le même programme modifié finit en 1962 par battre le champion du Connecticut, R. W. Nealey. Cf. **D. Michie**, *On Machine Intelligence*, op. cit., p. 39 qui cite le rapport publié dans le IBM Research News.

Un point mort peut être soit intérieur soit extérieur à la limite de profondeur fixée à l'avance; dans ce dernier cas, il n'a pas d'intérêt pour la poursuite de la stratégie optimale. Un point mort est donc un point soit objectivement inaccessible soit stratégiquement inaccessible : dans le premier cas, en modifiant la profondeur de la représentation virtuelle, on peut parvenir à l'atteindre; dans le deuxième, il est sans intérêt de l'atteindre, même à cette profondeur de jeu.

La notion de point mort fait donc partie de celle de '"stratégie compressée" visant l'abstraction et la généralisation. Un dernier axe de recherche va dans le même sens : celui de l'apprentissage des machines.

223. L'apprentissage des machines

Turing a montré que les machines étaient capables d'apprentissage si l'on faisait varier les poids d'évidence qui sont attachés aux différents critères adoptés pour la stratégie. Bien que l'idée d'apprentissage soit de Turing, elle n'a pas été développée directement par lui mais plus tard, à partir du début des années 60[216]. Faute de moyens techniques, en particulier touchant les capacités de stockage d'informations en mémoire, Turing n'a pas pu mettre véritablement en pratique l'idée d'un apprentissage de la part des machines, avec les ordinateurs dont il disposait au début des années 50.

Il faut étudier encore un cas de jeu, outre ceux des puzzles et des jeux à plusieurs, avant d'en venir à la caractérisation du jeu de l'imitation : c'est le cas des jeux que l'on pourrait appeler "intermédiaires", dans la mesure où ils semblent se situer à mi-chemin entre les deux premiers cas. Ce dernier cas, bien qu'il soit ambigu, possède une importance réelle dans la mesure où, comme on l'a vu, c'est précisément le cas du jeu de l'imitation que de se situer à mi-chemin entre puzzle et jeu.

23. Les jeux "intermédiaires"

---

[216] Cf. Cf. **D. Michie**, *On Machine Intelligence*, op. cit., p. 39-42. D. Michie concluait encore en 1974 que, même pour les jeux les plus simples, il n'y avait pas d'algorithme d'apprentissage qui garantirait un apprentissage optimal.

Turing a inventé plusieurs sortes de jeux qui ont servi à la constitution du jeu de l'imitation. Deux semblent particulièrement importants pour notre propos.

231. Le jeu de "Presents"

A. Hodges, dans sa biographie de Turing, fait allusion à un jeu inventé par Turing vers 1949 qui s'appelle "Presents" et dont il décrit brièvement les règles. Celles-ci introduisent pour la première fois un jeu à trois personnes qui ressemble au jeu de l'imitation [217]:

> **«L'idée en était qu'une personne sortait de la pièce et que les autres dressaient une liste de cadeaux imaginaires qu'ils pensaient qu'il souhaiterait recevoir. Ensuite, il revenait et pouvait poser des questions au sujet des cadeaux avant de les choisir et c'est là que le jeu de bluff et de double bluff commençait, car l'un des cadeaux était secrètement appelé "Tommy" et une fois que Tommy était choisi, la partie était terminée».**

Remarquons que dans ce jeu, les deux joueurs qui restent dans la pièce et qui dressent la liste des cadeaux comprenant le cadeau caché ne sont pas en situation de concurrence mais de connivence. Il s'agit d'une différence capitale par rapport au jeu de l'imitation. Le jeu ne se situe donc pas entre ces deux joueurs mais entre l'interrogateur et eux. Or, du point de vue de ce dernier, le jeu se ramène en fait à un puzzle, dans la mesure où il doit réussir à énumérer une liste de cadeaux conformes à ses désirs : l'énumération d'une liste peut être opérée par un algorithme, si tant est qu'une liste dressée à partir d'un désir puisse être l'objet d'une mise en algorithme. C'est d'ailleurs le rôle attribué au cadeau dénommé "Tommy" que de déjouer la pure et simple application d'un algorithme et d'exiger de la part de l'interrogateur la mise au jour de ses désirs. De ce point de vue, le "Tommy" représente ce que les deux autres joueurs croient être un désir caché de l'interrogateur et il fait office de "point mort" dans la mesure où, si l'interrogateur ne peut ou ne veut pas l'avouer, il lui est bien inaccessible objectivement ou stratégiquement[218].

---

[217] Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 397. Turing jouait à ce jeu en compagnie de Nick Furbank et Robin Gandy.

[218] Le choix du mot "Tommy" vient peut-être de l'expression anglaise [*Peeping Tom*] qui désigne le voyeur : du point de vue des deux jouerus qui dressent la liste, c'est finalement à une attitude de "voyeur" que revient la constitution du cadeau "inavouable".

Un autre jeu semble avoir servi à la constitution du jeu de l'imitation : il s'agit du jeu de "Psychology".

232. Le jeu de "Psychology"

Il y a en effet un autre "ancêtre" au jeu de l'imitation, plus ancien que le jeu de "Presents" inventé juste avant le jeu de l'imitation, et dont Turing a étudié les issues par le biais du calcul des probabilités. Turing l'avait inventé aux États-Unis quand il faisait son PhD de logique à Princeton en 1937. Par la suite, Turing a continué à jouer à ce jeu et à en examiner les issues.

232. 1. Traduction de la première section du manuscrit de "Psychology"

Voici comment il se présente dans les manuscrits de Turing [219] :

« Joueurs A et B.
A chaque partie, on choisit tout d'abord un nombre entre 0 et 1. Fréquence uniforme. Chaque joueur choisit un nombre entre 0 et 1.
Si A choisit α et B choisit β, alors B paye A

$$\frac{1}{2} X \operatorname{sgn}(\alpha - \beta)$$

Quoiqu'il en soit, les joueurs sont obligés en moyenne de jouer des nombres 0 à 1 avec une fréquence uniforme.
Si A joue un nombre $(\alpha, \alpha + d\alpha)$ avec une fréquence $\phi(\alpha, X)\, d\alpha$ et B joue un nombre $(\beta, \beta + d\beta)$ avec une fréquence $\psi(\beta, X)\, d\beta$ alors le gain moyen pour A est

$$\int_0^1 \int_0^1 \int_0^1 X\, \phi(\alpha, X)\, \psi(\beta, X) \operatorname{sgn}(\alpha - \beta)\, d\alpha\, d\beta\, dX$$

et les fonctions $\phi(\alpha, X)$, $\psi(\beta, X)$ sont sujettes aux conditions suivantes :

$$\int_0^1 \phi(\alpha, X)\, dX = 1 \qquad \int_0^1 \psi(\beta, X)\, dX = 1$$

---

219 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 128 et 373. Andrew Hodges parle du jeu de "Pyschologie" mais n'en mentionne pas les règles. Ayant eu l'occasion de lui demander personnellement à quoi celui-ci ressemblait, il m'a conseillé de m'adresser à Robin Gandy qui a joué à ce jeu avec Turing. Ce dernier ne se rappelait plus précisément des règles mais il m'a fait parvenir des photocopies tirées des archives Turing décrivant le jeu. Sans être parfaitement explicites, elles m'ont permis de reconstituer au moins son allure générale.

$$\int_0^1 \phi\,(\alpha, X)\, dX = 1 \qquad\qquad \int_0^1 \psi\,(\beta, X)\, dX = 1$$

Nous faisons la conjecture euristique que ce jeu est équivalent à la variation suivante :
Les restrictions

$$\int_0^1 \phi\,(\alpha, X)\, dX = 1 \quad \int_0^1 \psi\,(\beta, X)\, dX = 1$$

sont relaxées, mais le gain est

$$\frac{1}{2} X \ \ \mathrm{sgn}\,(\alpha - \beta) - f(\alpha) + f(\beta)$$

pour une fonction f. On déterminera ultérieurement cette fonction de sorte que la meilleure stratégie pour la fonction f aura pour résultat la restriction mentionnée. Nous montrerons ensuite que la solution obtenue ainsi donne la meilleure stratégie pour le jeu originel.»

232. 2. Remarques sur le jeu de “Psychology”

Il s’agit d’un jeu binaire de somme nulle, c’est-à-dire d’un jeu joué par deux joueurs A et B faisant des choix à partir d’un ensemble d’options, indexées sur l’intervalle réel [0, 1]. Les choix définissent une fonction de probabilités. Le choix de α par A et de β par B donne lieu à un paiement $K_{\alpha\beta}$ au joueur A et $-K_{\alpha\beta}$ au joueur B : il s’agit d’un jeu à somme nulle puisque la quantité $|\, K_{\alpha\beta}\, |$ gagnée par un joueur est perdue par l’autre. Le gain est alors défini par une formule du type : $K = 1/2\ \mathrm{sgn}\,(\alpha - \beta)$. Si $K > 0$, B paye A et si $K < 0$, A paye B. On peut faire deux remarques sur le déroulement du jeu.

Premièrement, pour toutes les parties, A doit jouer α avec une probabilité de gain $d\alpha$. La condition $\phi\,(\alpha, X)$ est la probabilité pour que A joue α dans le jeu X. La condition $\int_0^1 \phi\,(\alpha, X)\, dX = 1$ veut dire que, sur toutes les parties, A doit jouer α avec une probabilité $d\alpha$ dans n’importe quel intervalle $[\alpha, \alpha + d\alpha]$. Sans cette condition, A jouerait toujours $\alpha = 1$ et maximiserait ainsi ses gains : il y aurait là un “algorithme” évident pour gagner la partie et dans ce cas, le jeu se transformerait immédiatement en puzzle. Comme le fait remarquer R. Gandy [220]:

> **«Mais cela fait que le “jeu” n’en est plus un au sens habituel du terme - comment B pourrait-il dire que A n’était pas un en train de tricher en jouant toujours 1. Pour “jouer”, A choisit une fonction f (a, X), B choisit y (b, X) et ils calculent ensuite le gain à l’avance (ce que Turing appelle le “gain moyen”)».**

Le jeu n’est donc pas un jeu véritable dans la mesure où il n’impose pas de règles concernant le bluff : *c’est cette absence de règles touchant le bluff qui transforme le jeu en puzzle*.

Deuxièmement, on calcule les gains à l’avance en définissant une fonction représentant un optimum de gain. Comment définir cet optimum ? Pour étudier celui-ci, Turing émet ce qu’il appelle une “conjecture euristique” qui consiste à déplacer le cas étudié vers un cas qu’il juge plus simple et qui peut servir d’optimum dans le cas du jeu originel. Cette hypothèse consiste à modifier la fonction de distribution de X, ce qui renforce l’un des deux joueurs en favorisant la croissance de f en α. C’est ce qui permet ensuite de définir un optimum dans la conjecture euristique[221] puis dans la version originelle du jeu[222]. Le jeu de Psychologie se dédouble donc comme le jeu de l’imitation : la conjecture euristique émise par Turing consiste à inventer comme un “jeu n°2” dans lequel la recherche d’un optimum soit plus facile à définir que dans le cas premier. Mais, dans le jeu de l’imitation, le problème n’est pas, comme dans le jeu de

---

[220] Correspondance personnelle du 10 juin 1993.

[221] Turing définit une solution de type continu. Il aboutit à la conclusion - conjecturale, comme il le précise lui-même - que la meilleure stratégie pour A est :

$$\phi(\alpha, X) = \frac{1}{-X \log f(\alpha)} \qquad \text{où } \alpha = f(\alpha) \log e/f(\alpha)$$
$$\text{à condition que } 0 < \alpha < X \log e/X$$
$$= 0 \text{ autrement}$$

R. Gandy envisage un type discret de solution en faisant usage de programmation linéaire et fait remarquer qu’il n’a aucune idée de la façon dont Turing est parvenu à définir sa stratégie propre. Il ajoute cependant que Turing a peut-être constitué sa solution en l’adaptant à partir du cas continu.

[222] Le retour au jeu originel et la définition d’un optimum dans ce cas m’a paru incompréhensible. R. Gandy qui s’est essayé à la reconstituer avoue lui-même ne pas comprendre comment Turing est parvenu à cette solution.

Psychologie, de rapporter le cas n°2 au cas n°1 mais *l'inverse* : en effet, selon la thèse de Turing qui décrit le passage de l'informel au formel, c'est le passage de la première à la deuxième formule du jeu qui est problématique. Bref, on ne peut pas constituer le jeu n°2 en conjecture euristique ayant le même statut que dans le jeu de Psychologie tant que l'on n'a pas la preuve que le jeu n°2 remplace sans pertes le jeu n°1. Or cette preuve doit précisément être apportée par le déroulement d'une partie. Il y aurait donc un cercle vicieux à vouloir présupposer *a priori* la réussite du passage de la première à la deuxième formule du jeu et à accorder à sa deuxième formule un statut de conjecture euristique semblable à celui du jeu de Psychologie. C'est ce qui distingue profondément les deux jeux.

Le jeu de Psychologie apparaît donc bien comme un “ancêtre” du jeu de l'imitation, de par sa structure et ses conditions de jeu[223]. Néanmoins, le jeu de l'imitation, dans la mesure même où il repose sur la clarification de la différence ente le physique et l'intellectuel se distingue du jeu de Psychologie qui porte sur la détermination du choix d'un nombre réel dont la détermination est tout *abstraite*. En déplaçant son point de vue et en le faisant porter sur la différence entre le physique et l'intellectuel, Turing change de perspective et transforme du même coup la portée de l'argumentation probabiliste du jeu de l'imitation.

233. Retour au jeu de l'imitation

Comme on sait, c'est un jeu qui se joue à trois, un interrogateur et deux joueurs dont Turing suppose qu'ils sont en situation de concurrence (on pourrait en effet imaginer qu'ils se liguent contre l'interrogateur pour que celui-ci n'opère pas les bonnes identifications). Le jeu de l'imitation oscille entre la structure de jeu et la structure de puzzle selon que l'on se place du point de vue des joueurs ou de l'interrogateur d'une part et selon que l'on accorde une issue positive ou négative au jeu d'autre part. Étudions ces différents points de vue.

Les deux joueurs, homme et femme, sont en situation de concurrence sans véritablement jouer entre eux : leurs réponses peuvent gêner l'autre joueur mais

---

[223] Pourquoi porte-t-il ce nom ? Selon notre interprétation, c'est son aspect intermédiaire entre la notion de puzzle et celle de jeu qui lui accorde une importance du point de vue de la modélisation de la psychologie. c'est peut-être pour cette raison qu'il s'appelle ainsi.

elles ne visent pas à le vaincre puisqu'elles font essentiellement partie du jeu qu'ils jouent avec l'interrogateur.

L'interrogateur se trouve au départ confronté à deux adversaires dans ce qu'il croit être un jeu : il lui faut, comme les joueurs, appliquer une stratégie, se souvenir des réponses qui lui ont été faites et qui peuvent l'aiguiller vers des questions embarrassantes pour les joueurs. On voit que les positions de l'interrogateur et des joueurs sont inversées par rapport au jeu de "Presents" : alors que dans "Presents", c'est l'interrogateur qui peut avoir du mal à avouer l'objet d'un désir que les joueurs ont imaginé pour lui, dans le jeu de l'imitation au contraire le but - et non le désir - de l'interrogateur est clair (c'est l'établissement de la différence des sexes) mais les joueurs font tout ce qui est verbalement possible pour qu'il ne puisse pas y avoir accès. Bref, dans le jeu de l'imitation, l'interrogateur n'a pas à effectuer personnellement un travail sur ce qu'il peut chercher à ignorer de lui-même : il n'a pas de désir inconscient. Mais, du point de vue de l'observateur extérieur, on pourrait dire que les joueurs représentent le travail en question puisque les joueurs tentent autant que faire se peut de le maintenir dans l'illusion. Plusieurs cas de figure peuvent se présenter alors, selon que l'interrogateur parvient ou croit parvenir à opérer les bonnes identifications. Les deux premiers cas sont ceux pour lesquels l'interrogateur parvient à identifier les joueurs; le dernier est celui pour lesquels il n'y parvient pas.

Premièrement, dans le cas où l'interrogateur parvient à identifier un joueur dans le jeu n°1, c'est-à-dire dans le cas où l'identification se fait suffisamment vite pour que l'observateur extérieur n'ait pas le temps de remplacer l'homme par un ordinateur, la différence des sexes est déterminée et la liste des questions que l'interrogateur a posée lui paraît comme autant d'étapes dans le déroulement d'un algorithme. Le jeu de l'imitation se réduit alors à un puzzle pour tous les participants au jeu, y compris l'observateur extérieur.

Deuxièmement, si une partie dure suffisamment pour que l'observateur extérieur ait le temps de remplacer l'homme par un ordinateur et que l'interrogateur parvient à identifier un joueur dans le jeu n°2, deux cas sont possibles : soit il a identifié la femme, soit la machine.

Dans le cas où l'interrogateur a identifié la femme, la différence des sexes lui *paraît* être déterminée et il lui semble que le déroulement d'une partie est identique au déroulement des étapes d'un algorithme. Il croit donc que le fait d'avoir découvert qui était la femme lui permet de déterminer son complémentaire, l'homme, *mais il se trompe*. Dans la mesure où l'on se trouve dans un cas où l'interrogateur croit avoir réduit le jeu à un puzzle, on peut, par analogie avec la situation qui existe pour les algorithmes dans les systèmes formels, dire qu'il se trouve dans le cas de la détermination d'un ensemble récursivement énumérable non récursif : la liste des questions qu'il a posées ne permet pas la détermination indirecte du complémentaire de l'ensemble à déterminer. Mais cette analogie formaliste ne vaut que parce que l'interrogateur est dans l'erreur et même dans l'illusion quant à la nature de la différence des sexes. En fait, du point de vue de l'observateur extérieur, l'interrogateur croit *pour de mauvaises raisons* qu'il a réduit le jeu de l'imitation à un puzzle, même si le jeu se réduit effectivement à un puzzle.

Dans le cas où l'interrogateur a identifié la machine, il a réussi à s'apercevoir que l'observateur extérieur a changé les règles du jeu sans l'avertir. La différence des sexes est déterminée et joue le rôle d'un critère de jeu comme un autre, que l'algorithme mis en place par l'interrogateur est susceptible de dépasser. Le jeu se réduit à un puzzle pour l'interrogateur comme pour l'observateur extérieur.

Troisièmement, dans le cas où le jeu dure et que l'interrogateur ne parvient à identifier un joueur ni au jeu n°1 ni au jeu n°2, la différence des sexes reste pour lui indéterminée et sa liste de questions entre dans le cadre d'une stratégie de jeu. Qu'en est-il pour l'observateur extérieur ? S'agit-il d'un jeu ou d'un puzzle ? Il s'agit en fait d'une question de point de vue. L'observateur extérieur doit en effet reconnaître qu'il s'agit d'un *jeu* à l'intérieur pour l'interrogateur et d'un *puzzle* à l'extérieur pour lui. Il doit en effet prendre en compte le fait que l'interrogateur est dans l'illusion et le fait que lui ne l'est pas : la différence des sexes existe et n'existe pas selon que l'on se place à l'intérieur ou à l'extérieur du jeu.

C'est l'observateur extérieur qui est censé réduire le jeu à un puzzle : si c'est bien le cas, le jeu n°2 jouerait le même rôle que celui d'un puzzle de substitution. Cette réduction est-elle possible ? Si l'on reprend les propres termes de l'analyse des jeux par Turing, il faudrait réussir à affecter des poids d'évidence aux réponses dans le jeu n°1 et montrer qu'en maintenant un aspect ambigu dans celles-ci, il est possible d'évoluer du jeu n°1 au jeu n°2 sans solution de continuité. Peut-on affecter des poids d'évidence tels que la différence entre le physique et l'intellectuel ne soit jamais assignable par l'interrogateur ? Comment réussir à affecter d'une valeur numérique sur une échelle d'évaluation les réponses des joueurs sans préjuger à l'avance de l'issue générale d'une partie ou même de ses issues partielles ? Cette tâche serait pourtant indispensable à qui voudrait constituer une évaluation en termes de minimax. En particulier, on ne voit pas où constituer des "points morts" au-delà desquels il n'y aurait pas intérêt à explorer les solutions pour établir une différence radicale entre le physique et l'intellectuel : comme le nombre de réponses possibles pour une question donnée est énorme, une évaluation en termes de minimax n'est plus praticable.

Cependant, si l'on abandonne, faute de moyens, la poursuite de la *justification* de ce troisième cas de figure et si l'on fait l'hypothèse que la solution de Turing est la bonne, on peut réussir à *interpréter* la solution adoptée par Turing et saisir ses intentions profondes.

3. Hypothèse de la réussite de l'argument probabiliste

Plaçons-nous donc au moment-limite de 2050 et tâchons de voir ce qui se produirait si l'on accordait à Turing la pleine réussite de l'expérience du jeu de l'imitation, c'est-à-dire si l'on faisait l'hypothèse qu'il s'agit bien d'un jeu à l'intérieur et d'un puzzle à l'extérieur.

31. Probabilités et algorithmes

L'interprétation probabiliste voit dans les comportements des participants au jeu un rapport de cryptage et de décryptage des messages transmis : les joueurs introduisent volontairement du bruit parasite dans leur réponse et l'interrogateur essaye d'éliminer ce bruit. Le bruit dans le message apparaît donc comme *la trace de l'information portant sur la nature physique des joueurs* : en effet, du point de vue de l'interrogateur, les réponses formulées par les joueurs demandent à être décryptées puisqu'elle tentent de cacher toute information qui le mettrait sur la piste d'une identification de la nature physique des joueurs; du point de vue des joueurs, les réponses qu'ils formulent demandent à être cryptées puisqu'il s'agit de limiter autant que possible tout ce qui pourrait mettre l'interrogateur sur la piste d'une identification de leur nature physique.

Du point de vue de l'observateur extérieur, l'interrogateur n'a plus aucune chance d'opérer les bonnes identifications. Le cryptage des informations peut, sans aucun risque, être effectué aussi bien par un homme que par un ordinateur. Contrairement à ce qui était apparu dans l'interprétation formaliste, l'interrogateur apparaît dans ce contexte comme irrémédiablement humain, c'est-à-dire comme ne pouvant jamais avoir accès à l'aspect universel du concept d'intelligence. C'est pourquoi la différence entre le point de vue interne et le point de vue externe n'est, pour l'interrogateur, jamais abolie.

Du point de vue de l'argumentation générale, comme on l'a déjà souligné, parvenir en esprit en 2050 implique que le lecteur-arbitre reconnaisse l'identité du comportement humain et du comportement de l'ordinateur d'une part et s'identifie lui-même à une machine d'autre part. Ce que le lecteur-arbitre conçoit et que l'interrogateur ne peut pas concevoir, c'est que l'ordinateur peut non seulement remplacer l'homme dans le jeu mais que le jeu entier peut être interprété d'un point de vue extérieur comme un *rapport de cryptage / décryptage entre algorithmes effectués par des ordinateurs digitaux*.

Comment caractériser le bruit de ce point de vue ? Du point de vue des algorithmes effectués, le bruit se manifeste alors sous l'apparence d'un aléatoire qu'il faut tenter de réduire quand on se place du point de vue de l'interrogateur et qu'il faut engendrer quand on se situe du point de vue des joueurs. On doit donc considérer que ce qui se manifeste du point de vue algorithmique comme aléatoire est seulement une *absence* d'information et non pas la trace de l'information portant sur la nature physique des joueurs.

Deux conceptions de l'aléatoire s'affrontent donc dans le jeu, selon que l'on se place du point de vue de l'interrogateur ou de celui du lecteur-arbitre. Du point de vue de l'interrogateur, c'est-à-dire dans ce qu'il considère comme un jeu, l'aléatoire est une trace du physique, tandis que du point de vue du lecteur-arbitre, c'est-à-dire dans ce qu'il considère comme un puzzle, l'aléatoire n'est qu'une absence d'information. Comment, à partir de la même notion d'algorithme, parvenir à deux conceptions aussi différentes de l'aléatoire ?

32. Aspects intellectuels et physiques de la notion d'algorithme

Jusqu'à présent, ce qui dépassait le cadre du calcul algorithmique nous était apparu, dans le contexte du problème de l'arrêt, sous le jour de la question logique de la décision. Ici, ce qui dépasse le cadre en question est lié à la notion d'aléatoire dans un contexte physique. Y-a-t-il un rapport entre ces différentes déterminations ?

32.1. Les algorithmes, la physique et l'aléatoire

D'un point de vue épistémologique, on peut caractériser par deux traits la nature de la notion d'algorithme. Il est en effet possible de distinguer un aspect logique et un aspect physique dans celle-ci puisqu'il est possible d'interpréter un dispositif *physique* comme manifestant la présence *idéale* d'un algorithme, sans les confondre : c'est le propre de l'ordinateur digital en tant que machine universelle incarnée de pouvoir effectuer plusieurs tables d'instructions. Insistons sur ces deux aspects, logique et physique.

D'une part, un algorithme permet de considérer une classe infinie de questions comme pur objet intellectuel quel que soit le nombre des éléments de la classe, puisque le rôle de l'algorithme est d'établir une distinction radicale entre la nature potentiellement infinie de la classe de questions et le caractère fini de la règle qui permet de répondre à chacune. On retrouve ici la différence qui existe entre la position occupée par le lecteur-arbitre en possession de la règle finie d'engendrement d'un algorithme et l'interrogateur qui ne possède pas cette règle finie mais seulement un stock indéfini de questions nécessaires pour tenter de limiter l'aspect aléatoire des réponses reçues.

D'autre part, c'est parce que la règle définissant l'algorithme est de nature finie qu'il est possible d'envisager l'algorithme sous l'aspect d'une machine de Turing et d'incarner cette machine abstraite dans une machinerie physique. Si l'algorithme n'était pas de nature finie, c'est-à-dire si les instructions susceptibles d'en rendre compte occupaient un nombre infini de cases sur le ruban d'une machine, l'incarnation du ruban dans une machinerie matérielle ne pourrait pas avoir lieu.

C'est précisément sur ces deux aspects que porte le jeu de l'imitation. C'est dans le rapport entre ces deux caractéristiques, logique et physique, que se manifeste la présence de l'aléatoire.

32. 2. L'aléatoire logique

Quand on se trouve dans la position de l'interrogateur du jeu de l'imitation, le meilleur moyen de déchiffrer un message chiffré consiste à posséder les algorithmes d'encryptage, qu'il a fallu réussir à reconstruire en se laissant guider par le sens supposé que l'on prête aux messages. C'est là une difficulté majeure car il peut être extrêmement difficile, voire impossible, de reconstituer la règle d'encodage qui a produit un message chiffré.

Prenons l'exemple du nombre π, dont Turing a démontré dans "On Computable Numbers…"[224] qu'il faisait partie des réels calculables en exhibant un algorithme le calculant. La règle d'engendrement du nombre π choisie par Turing est $\pi = 4\ (1 - 1/3 + 1/5 - \ldots)$; en elle-même, cette règle ne dit en rien pourquoi, après le premier demi-million de décimales, on trouve neuf décimales formant la suite naturelle des chiffres : 123456789. Les caractéristiques particulières de l'objet échappent ainsi à ses conditions d'engendrement parce

---

[224] **A. M. Turing**, (1936), "On Computable Numbers (…)" § 10, republié dans [**M. Davis ed.**, *The Undecidable*, op. cit.], p. 140.

qu'elles sont isolées les unes des autres. Ces caractéristiques apparaissent sous l'aspect d'un message chiffré dont on ne posséderait pas la clé.

Dès lors, dans le cas où l'on a seulement accès à certaines caractéristiques d'un nombre réel sans connaître sa règle d'engendrement, peut-on réussir à partir des caractéristiques que l'on possède à *remonter* à la règle d'engendrement du nombre en question ? Par exemple, si l'on possède la liste des décimales de π après les quatre premiers millions de chiffres, sans savoir qu'il s'agit des décimales de π, est-on capable de *reconnaître* qu'il s'agit bien des décimales de π après les quatre premiers millions de chiffres ? Non, sauf à avoir déjà calculé ces quatre millions de chiffres, si l'on a eu auparavant l'idée de penser qu'il pouvait s'agir des décimales de π, alors que rien ne le laisse supposer.

Aussi la règle d'extraction des décimales de π déploie-t-elle nécessairement ses décimales mais cette nécessité nous échappe-t-elle complètement si l'on ne possède que le résultat de la règle d'extraction, c'est-à-dire les décimales en question : il est en effet absolument impossible de reconstituer la règle et encore moins de retrouver de quel nombre il s'agit quand on se retrouve confronté à une suite de chiffres qui paraît aléatoire.

Dans quel contexte et sous quel type de description cet aspect aléatoire peut-il revêtir une signification physique ?

32. 3. L'aléatoire physique

Pour comprendre en quel sens l'aléatoire algorithmique peut avoir un sens physique, Turing fait remarquer l'existence d'une parenté de point de vue entre deux sciences : la cryptographie et la physique.

32. 31. La cryptographie et la physique

Cette comparaison prend la forme suivante [225]:

> **«Il y a un parallèle remarquablement proche entre les problèmes du physicien et ceux du cryptographe. Le système dans lequel un message est codé correspond aux lois de l'univers, les messages interceptés à ce qui est manifeste, les clés pour la journée ou un**

---

[225] **A. M. Turing**, “Intelligent Machinery”, op. cit., p. 40

message à des constantes importantes qui doivent être déterminées. La correspondance est très proche, mais l'objet de la cryptographie est très facilement traité avec une mécanique discrète, ce qui n'est pas si facile dans le cas de la physique.»

Le parallèle entre les deux domaines rend compte de la situation dans laquelle se trouvent l'interrogateur et l'observateur dans le jeu de l'imitation. En effet, ce que l'interrogateur tente de reconstituer à partir de messages codés, c'est le signe *abstrait* d'une différence *physique*. Si ce signe n'existe pas, il n'est pas possible de retrouver la différence physique qui existe entre les joueurs, différence qu'ils cryptent dans leurs réponses. Dans ce cas de figure, l'aspect aléatoire indique non pas l'absence d'une différence physique réelle mais la présence d'une différence physique inassignable. L'aléatoire a donc ici un sens physique.

Du point de vue de l'observateur extérieur, cette différence physique réelle mais inassignable peut dès lors être considérée comme négligeable puisque l'interrogateur n'a pas les moyens d'en prendre conscience à l'intérieur du jeu. L'aléatoire a ici un sens seulement logique.

Si l'on suit le parallèle entre la cryptographie des questions et des réponses au jeu de l'imitation et la physique, en quel sens cette dernière doit-elle être considérée comme déterministe ? La question se pose puisque l'aléatoire, du point de vue de l'interrogateur, manifeste la différence physique entre les joueurs et possède bien de ce fait un fondement physique tout en étant inassignable tandis que ce même aléatoire, du point de vue de l'observateur extérieur, brouille les messages sans avoir de fondement physique. Deux conceptions différentes de la physique sont engagées par ces deux attitudes puisque l'aléatoire n'y possède pas la même signification : dans la première, la physique est déterministe puisque, idéalement, la réduction totale de l'aléatoire permettrait d'assigner une signification physique à tout message; dans la seconde, la différence physique entre les joueurs peut toujours être établie, mais sans que l'aléatoire transmette aucune information ayant trait à la nature physique puisqu'il a un statut exclusivement logique. Comme on sait par ailleurs que l'interrogateur a l'illusion de croire qu'il pourra parvenir à réduire l'aléatoire mais qu'il y parviendra de moins en moins, sa conception déterministe de la place de l'aléatoire en physique est, elle aussi, une illusion. Le résultat du jeu conduit donc à concevoir la physique comme le fait l'observateur extérieur : elle est déterministe et la présence de l'aléatoire ne manifeste, d'un point de vue logique, que notre ignorance. De ce point de vue, l'aléatoire que l'interrogateur conçoit comme le contraire de l'information, c'est-à-dire comme du *bruit* qu'il faut réussir à éliminer, est conçu par l'observateur comme de l'information *redondante* et par conséquent éliminable.

32. 32. Discret et continu du point de vue algorithmique

Quel que soit ce que l'on pense du résultat ultime du jeu dans l'interprétation probabiliste, la mise en parallèle de la cryptographie et de la physique peut cependant sembler étrange dans la mesure où Turing lui-même souligne plus les différences qui existent entre les deux disciplines que leur ressemblance.

Comme l'indique la dernière phrase de la citation, la différence entre la cryptographie et la physique vient de ce que la première situe son domaine d'intelligibilité dans un univers *discret* tandis que la seconde se situe dans un univers *continu* : c'est ce qui explique qu'il ne soit «pas si facile» d'en faire une modélisation discrète par le biais de la notion d'algorithme. Turing ne s'explique pas sur les raisons qui le poussent à considérer la physique comme relevant plutôt du continu, mais on peut comprendre pourquoi si l'on remarque que la physique fait, dans ses modèles, usage des nombres réels, à la fois pour des raisons géométriques liées à la représentation de l'espace et du temps et pour des raisons algébriques liées à l'utilisation d'équations différentielles[226]. Dans ce contexte, l'articulation des aspects continus et discrets de l'univers physique revêt une importance fondamentale. Or l'utilisation du calcul des probabilités permet précisément une meilleure articulation de ces aspects parce qu'il facilite le

---

[226] Cf. **R. Penrose**, *The Emperor's New Mind*, Oxford University Press, Oxford, 1989, p. 113 : «Le système des nombres réels est choisi en physique pour son utilité *mathématique*, sa simplicité, son élégance, en plus du fait qu'il s'accorde, sur une large échelle, avec les concepts physiques de distance et de temps».

passage de l'un à l'autre[227] et c'est sans doute pourquoi Turing l'utilise dans le jeu.

Remarquons que c'est précisément cette articulation du discret et du continu mathématique qui faisait déjà l'objet de "On Computable Numbers …" puisque Turing y posait la question de la calculabilité des nombres réels. La filiation entre "On Computable Numbers …" et "Computing Machinery and Intelligence" se situe donc dans le rapport du continu mathématique et du continu physique à la question de la calculabilité, c'est-à-dire dans la question de la maîtrise des nombres réels, et non pas dans l'aspect formel le plus visible, à savoir l'utilisation des machines de Turing. On a donc ici une analogie très forte avec la perspective adoptée dans "On Computable Numbers …" : de même que les algorithmes des machines de Turing calculent des nombres réels, de même ici les rapports entre algorithmes permettent d'approcher la matière physique continue[228].

32. 4. L'aléatoire et le jeu de l'imitation

Si l'on en revient au jeu de l'imitation, on voit qu'il y a une analogie entre la façon dont Turing construit une interprétation probabiliste du jeu et la façon dont il décrit le rapport qui existe entre la notion d'algorithme et de science physique.

En effet, la matière continue joue dans le rapport algorithme / physique le rôle de terme extrême qui se manifeste par le biais de l'aléatoire, de même que la femme joue le rôle du terme extrême dans les deux formules du jeu : adversaire de l'homme, elle devient l'adversaire de l'ordinateur et le rapprochement entre l'homme et l'ordinateur se fait par opposition à elle. En poursuivant l'analogie, on peut donc dire que la matière physique possède, en tant que matière, deux aspects

---

227 Comme le faisait remarquer E. Borel : «Le calcul des probabilités apparaît ainsi comme l'instrument mathématique extrêmement souple qui permet, tantôt de passer du continu au discontinu, tantôt au contraire d'introduire le continu dans l'étude de certains phénomènes discontinus». **E. Borel**, Allocution à l'Académie des Sciences, séance du 17 décembre 1934, reproduit dans *Borel, philosophe et homme d'action*, Gauthiers-Villars, Paris, 1967, p. 364.

228 On peut dire de ce point de vue, en utilisant un vocabulaire kantien, que "Computing Machinery and Intelligence" joue le rôle d'une contrepartie "transcendantale" à l'analyse "formelle" de "On Computable Numbers …".

: d’une part un aspect discret et calculable qui est celui qui s’apparente à la calculabilité des réels et qui de ce fait autorise le remplacement de l’homme par l’ordinateur au jeu de l’imitation et d’autre part un aspect continu, aléatoire et non-calculable, qui reste le propre de la femme et d’elle seule au jeu de l’imitation. De ce point de vue, la femme remplit un rôle identique au continu matériel : sa présence dans le jeu se manifeste par le biais d’un aléatoire que l’interrogateur n’arrivera jamais à éliminer. Ainsi est-ce la différence entre le continu et le discret qui constitue la différence entre le puzzle et le jeu dans l’interprétation du jeu de l’imitation. L’aléatoire dans le jeu manifeste donc la présence réelle et cachée de la femme. Il faut donc en conclure que, pour Turing, *la femme est plus matérielle que l’homme*. Voici donc comment Turing parvient à articuler une conception logique et une conception physique de l’aléatoire : c’est en fait la différence des sexes qui, dans le cadre du jeu, permet cette articulation. Cette conclusion retrouve ce qui nous était déjà apparu à la fin du chapitre précédent : il y a bien une différence entre l’homme et la femme qui n’est pas abolie dans le jeu de l’imitation et qui empêche que soit pris directement en compte le genre humain en général.

On peut donc maintenant compléter le réseau sémantique mis en place par Turing : l’homme se range du côté du digital et de l’aspect déterministe de la notion idéale d’algorithme; c’est pourquoi il peut être remplacé par un ordinateur qui possède les mêmes caractéristiques. En revanche, la femme s’oppose à l’homme et à l’ordinateur parce qu’elle renvoie à l’aspect inaccessible de la matière, à savoir au continu et à sa manifestation intellectuelle : l’aléatoire.

Comment Turing a-t-il pu en venir à cette conception étrange de la place de la différence des sexes nécessaire - bien que sous-jacente - à la bonne marche du jeu de l'imitation ? C’est ce qu’il faut voir maintenant en étudiant les stratégies adoptées par les joueurs.

Chapitre IV

La viabilité du jeu de l'imitation

Si l'on note les différents indices donnés par Turing quant à la stratégie suivie par les joueurs, on s'aperçoit qu'il introduit des différences stratégiques importantes entre l'homme et la femme. C'est en introduisant ces différences que Turing tente de parvenir au but qu'il s'est fixé : montrer qu'il n'est pas possible pour l'interrogateur de réussir à deviner la nature physique des joueurs, que ceux-ci soient des êtres humains ou des ordinateurs digitaux. Ce sont donc les stratégies intellectuelles des joueurs dans leur rapport à leur nature physique qu'il s'agit d'étudier maintenant.

1. Les stratégies intellectuelles des joueurs

Différentes stratégies sont concevables pour les joueurs interrogés, en vue de cacher leur identité sexuelle ou leur absence d'identité sexuelle. Il nous faut donc maintenant étudier la notion de bluff.

11. La stratégie de la femme : "être femme"

Turing précise que la stratégie de la femme doit être de dire la vérité, sans donner clairement de raisons à ce choix. Turing donne un exemple de stratégie de la femme [229] :

> **«La meilleure stratégie pour elle est sans doute de donner des réponses vraies. Elle peut ajouter à ses réponses des choses comme "C'est moi la femme, ne l'écoutez pas !", mais cela ne mènera**

[229] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 434.

à rien puisque l'homme peut faire des remarques semblables».

La femme redouble donc l'affirmation de son identité sans chercher à la cacher. La stratégie optimale de l'interrogateur revient donc finalement à réussir à identifier la femme et non pas l'homme une fois qu'il s'est rendu compte qu'elle n'utilisait pas de bluff. Comme le fait remarquer A. Hodges [230]:

«La façon si fameusement irrévérencieuse que Turing a de commencer son article consiste à décrire un "jeu de l'imitation" dans lequel un interrogateur pose des questions à deux personnes qu'il ne voit pas et décide qui est la femme sur la base des réponses écrites - en fait par le biais d'une communication téléimprimée - dans lesquelles l'homme et la femme prétendent à la fois être la femme».

Si en effet la femme ne bluffe pas, c'est elle que l'interrogateur va parvenir à identifier le plus vite. La stratégie de la femme consiste seulement à être elle-même, ce qui caractérise précisément l'absence de stratégie selon la règle du jeu. Mais pourquoi la femme devrait-elle dire la vérité ? Il est impossible de répondre à cette question au vu des indications données par Turing et il faut donc considérer qu'il s'agit d'un présupposé, à mettre en rapport avec ce que nous avons déjà appris des caractères attribués à la femme : son aspect "matériel" , qui relève du continu.

12. La stratégie de l'homme : "imiter la femme"

Pour brouiller l'interrogateur, l'homme doit tenter de se faire passer pour la femme[231]. Aussi, dans l'exemple donné par Turing, à la question de l'interrogateur concernant la longueur des cheveux de A, c'est-à-dire de l'homme, celui-ci répond comme s'il était une femme [232]:

«Mes cheveux sont coupés à la garçonne et les mèches les plus longues font à peu près vingt centimètres».

Le bluff de l'homme consiste donc à imiter la femme. Cette imitation peut-elle être effectuée avec succès par une machine ?

---

230 A. **Hodges**, (1988). *Alan Turing and The Turing Machine* in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988] pp. 9-10.

231 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 434.

232 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 434.

13. La stratégie de la machine : "imiter l'imitation de l'homme imitant la femme ou imiter la vérité de l'attitude de la femme"

Dans le jeu n°2 dans lequel la machine remplace l'homme, la machine doit, pour passer inaperçue, imiter la stratégie de l'homme imitant la femme. Turing donne trois exemples de cette stratégie qui ont tous les trois leur spécificité [233]:

> **«Question : Ecrivez-moi, s'il vous plaît, un sonnet au sujet du pont sur le Forth.**
> **Réponse : Ne comptez pas sur moi pour ça. Je n'ai jamais été capable d'écrire des poèmes.**
> **Q : Ajoutez 34 957 et 70 764.**
> **R : (En silence pendant 30 secondes puis donne la réponse) 105 621.**
> **Q : Jouez-vous aux échecs ?**
> **R : Oui.**
> **Q: J'ai mon roi en C8 et aucune autre pièce. Il vous reste seulement votre roi en C6 et une tour en A1. A vous de jouer. Que jouez-vous ?**
> **R : La tour en A8 , échec et mat.»**

Examinons les réponses données par la machine.

131. La première et la troisième réponse de la machine

La première réponse manifeste une incapacité de la part de la machine puisqu'elle admet ne pas pouvoir composer un poème. La manifestation de cette incapacité permet-elle de déterminer le sexe de celui ou celle qui fait cette réponse ? Il ne semble pas : beaucoup d'êtres humains seraient incapables d'écrire un poème au pied levé. La réponse est donc ambiguë et ne peut sans doute pas servir d'indice à l'interrogateur.

La troisième réponse de la machine, qui consiste à faire échec et mat à l'interrogateur manifeste très clairement les pouvoirs de la machine. Remarquons cependant que dans la configuration du jeu qui sert d'exemple à Turing, la solution apparaît d'elle-même : le nombre de solutions ayant une signification pour cette partie est si limité, vu le peu de pièces qui reste en jeu, que l'interrogateur ne peut pas deviner à qui il a affaire, un être humain ou une machine. On pourrait imaginer un exemple, comme celui du calcul arithmétique, où les pouvoirs de la machine sont tels que l'interrogateur finirait par se douter qu'il n'a affaire ni à un homme ni à une femme. C'est précisément ce cas que

---

[233] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., pp. 434-435.

Turing étudie dans la deuxième réponse.

132. La deuxième réponse de la machine

La deuxième réponse est très curieuse : le résultat donné par la machine est *faux*. Le résultat exact est 105 721 et non pas 105 621. S'agit-il seulement d'une coquille[234]? On peut émettre une autre hypothèse, plus vraisemblable. L'inexactitude du résultat n'est pas suffisamment grande pour que l'on ne puisse pas penser à une erreur d'inattention de la part de celui qui fournit la réponse : il s'agit visiblement d'une erreur dans l'addition de la colonne des centaines, où un 6 a remplacé le 7. Cette erreur peut donc passer pour une erreur d'inattention faite par un être humain, *quel que soit son sexe*.

Un passage de l'article corrobore d'ailleurs cette interprétation : Turing se fait à lui-même l'objection selon laquelle la machine serait immédiatement reconnue au jeu de l'imitation à cause de ses performances inhumaines en arithmétique[235] :

> **«On affirme que l'interrogateur pourrait distinguer la machine de l'homme en leur posant simplement un nombre de problèmes arithmétiques. La fatale exactitude de la machine la démasquerait ».**

C'est bien le caractère "fatal"[*deadly*] de l'exactitude de la machine en arithmétique qui lui enlève tout caractère humain. Aussi faire des erreurs dans les réponses aux questions est-il une nécessité pour faire croire que la machine possède une identité sexuelle : mentir en faisant des erreurs est la façon dont la machine acquiert une identité sexuelle au jeu de l'imitation. C'est pourquoi Turing peut dire [236]:

> **«La machine (programmée pour jouer au jeu) n'essayerait pas de donner les *bonnes* réponses aux problèmes arithmétiques. Elle essayerait délibérément d'introduire des**

---

[234] C'est ce que suppose la traduction française, dans laquelle le résultat a été pieusement corrigé en 105 721. Douglas Hofstadter avait, quant à lui, remarqué l'erreur dans la réponse Cf. **D. Hofstadter**, *Gödel,Escher et Bach*, Interéditions, Paris, 1985, p. 667-668. Hofstadter reste cependant peu précis sur la signification de cette "erreur" pour le jeu de l'imitation lui-même. Il remarque seulement : «La réflexion sur ce que Turing a pu vouloir dire au moyen de ce passage subtil soulève à peu près tous les grands problèmes philosophiques liés à l'intelligence artificielle.».

[235] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 448.

[236] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 448.

erreurs par un calcul destiné à déconcerter l'interrogateur».

Mais alors, la machine ne risque-t-elle pas d'être découverte par l'interrogateur, puisqu'elle n'imite plus la stratégie de l'homme qui lui-même imitait la femme ? Non répond Turing. Ce n'est plus l'homme spécifiquement que la machine imite parce qu'elle n'en a plus besoin dans le cas d'un problème arithmétique : les erreurs que la machine commet apparaissent comme la trace de la présence d'un *être humain* et non plus d'un homme ou d'une femme. L'interrogateur n'a pas les moyens de discerner dans ces erreurs autre chose que ce qu'elles sont, un calcul "trop" humain. Bref, la machine n'imite plus la capacité de travestissement de l'homme mais la capacité d'erreur de l'être humain, c'est-à-dire la *vérité* de l'être humain, qui est *par nature* mauvais en calcul. Quand on la compare à un ordinateur digital, cette incapacité humaine est tellement criante qu'elle en fait disparaître la différence des sexes. C'est donc par la manifestation de l'erreur humaine que la machine parvient à bluffer au jeu n°2.

Dès lors, la machine, dans le cas arithmétique, en revient à ce qui faisait la forme de la stratégie de la femme, à savoir son "honnêteté" : il s'agit pour la machine *d'être* humaine comme la femme *était* féminine sans chercher à imiter personne d'autre qu'elle-même. Mais dans ce cas, loin de se trouver éliminée par la règle du jeu, la machine, contrairement à la femme, occulte définitivement son absence d'identité sexuelle. C'est donc par la vérité que la femme était femme alors que c'est par l'erreur que la machine fait semblant d'être humaine. Il y a donc une autre stratégie pour la machine que d'imiter l'imitation de l'homme et ce, dans un cas très spécifique, celui dans lequel il y a à faire un calcul arithmétique. Cette autre stratégie est, cette fois, "féminine" puisqu'elle consiste à faire semblant de dévoiler son identité. Elle permet d'assurer définitivement que l'erreur de calcul n'est pas l'apanage d'un sexe parce que celle-ci ne relève pas de la différence sexuelle.

Un autre exemple semble corroborer cette analyse des stratégies de la machine. Turing, dans un exemple d'un dialogue entre l'interrogateur et un joueur[237] prend pour sujet de questionnement le sonnet de Shakespeare qui commence par : «Te comparerai-je à un jour d'été »[238]. L'interrogateur demande si "jour de printemps" ne serait pas meilleur. Le joueur répond que cela n'aurait pas le bon nombre de pieds [«*It wouldn't scan*»]. Il s'agit donc d'un jeu de mots sur le verbe [*scan*] (qui vient du verbe français "scander") qui renvoie à la fois à la "scansion" d'une poésie mais aussi à "l'observation" faite par une machine de Turing en face d'une case de son ruban infini. Il est à noter que le choix de ce poème de Shakespeare n'est pas gratuit : c'est Shakespeare que l'un des contradicteurs de Turing au sujet de la possibilité de l'intelligence artificielle, le professeur Jefferson, aimait à citer à l'appui de sa thèse contre la possibilité qu'une machine puisse apprécier la poésie[239], mais c'est aussi Shakespeare qui écrivait des poèmes d'amour dont le destinataire était vraisemblablement un *homme*, comme Turing le savait sûrement.

Comment la machine peut-elle parvenir à simuler l'erreur ?

133. Les simulations de la machine

133. 1. La simulation de l'erreur

La machine parvient à simuler l'erreur parce qu'il y a, dit Turing, deux types d'algorithme, c'est-à-dire deux types de programme, qu'une machine peut exécuter, quand on observe les "sorties" d'une machine donnée. Le premier type est de nature déterministe. Turing considère ce type d'algorithme comme une

---

[237] Cf. **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 446. Il ne s'agit pas exactement d'un dialogue du jeu de l'imitation puisque Turing dit expressément qu'il s'agit d'un jeu plus simple, joué entre un interrogateur et un seul joueur, la femme ayant été, comme Turing le précise, écartée. Ce qu'il appelle "jeu" est en fait un examen oral, appelé en jargon universitaire britannique un "viva voce" (l'expression étant souvent raccourcie en "viva") au cours duquel des questions sont posées à l'étudiant par un professeur. Turing a sans doute pensé à un examen universitaire parce qu'il répond à un professeur, le professeur Jefferson.

[238] Il s'agit du premier vers du sonnet XVIII, universellement connu en Angleterre.

[239] C'est en effet ainsi que le professeur Jefferson terminait sa "Conférence Lister de 1949", dont Turing cite un extrait dans "Computing Machinery and Intelligence", p. 445. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 405 qui cite le dernier paragraphe de cette conférence de Jefferson : « (…) me rangeant du même côté que l'humaniste Shakespeare plutôt que du côté des mécanistes, je rappellerai ces vers de Hamlet : "Quelle œuvre que l'homme ! Combien noble en raison ! Infini dans ses facultés (…)."».

manifestation du déterminisme laplacien [240]:

> **«Il apparaîtra qu'étant donné un état initial de la machine et des signaux d'entrée donnés, il est toujours possible de prévoir tous les états futurs. Cela rappelle le point de vue de Laplace selon lequel à partir de l'état complet de l'univers à un moment donné du temps, tel qu'il est décrit par les positions et les vitesses de toutes les particules, il devrait être possible de prévoir tous les états futurs».**

Ce programme aboutit, après un temps fini de calcul et si une solution existe *a priori*, à une réponse vraie. Ainsi dans la troisième réponse fournie par la machine concernant le jeu d'échecs, celle-ci résout le problème qui est posé dans la partie et fait échec et mat à l'interrogateur.

Le second type de programme produit des résultats qui apparaissent comme aléatoires. Il permet le traitement de solution sans que des considérations temporelles entrent en jeu puisqu'il n'exige pas de stockage en mémoire : les solutions déjà envisagées peuvent être reprises comme données. Turing en donne un exemple [241]:

> **«Un élément aléatoire est assez utile quand nous cherchons la solution d'un problème. Supposez par exemple que nous voulions trouver un nombre entre 50 et 200 qui soit égal au carré de la somme de ses chiffres. Nous pourrions commencer à 51, puis essayer 52 et poursuivre jusqu'à que nous obtenions un nombre qui soit satisfaisant. Ou bien nous pourrions choisir des nombres au hasard jusqu'à ce que nous en trouvions un bon. Cette méthode a l'avantage qu'il n'est pas nécessaire de garder trace des valeurs déjà essayées mais le désavantage qu'il est possible que l'on essaye deux fois la même, ce qui n'est pas très important s'il y a plusieurs solutions. […]. Puisqu'il y a probablement un très grand nombre de solutions satisfaisantes, la méthode aléatoire semble meilleure que la méthode systématique».**

Remarquons que cet exemple est, comme celui de l'addition des deux nombres, légèrement erroné, puisqu'il n'y a qu'un seul nombre entre 50 et 200 qui possède la propriété recherchée - le nombre 81 - et non pas plusieurs, comme le prétend Turing[242]. L'intervalle de nombres étant relativement restreint - entre 50 et 200 -, il est probable que la machine arrive à la solution en utilisant un algorithme aléatoire, bien que ce ne soit pas certain (la machine peut donc continuer indéfiniment à chercher une solution). Il y a donc bien possibilité, par le biais de l'aléatoire, de *faire croire* à une erreur.

---

240 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 440.
241 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 459.
242 En effet, on a 81 : $8 + 1 = 9$ et $9^2 = 81$.

133. 2. L’interprétation de l’erreur

Turing distingue deux types d’erreurs dans son article : les erreurs de fonctionnement et les erreurs de conclusion[243]. Dans le cas de la machine, les erreurs de fonctionnement sont dues à une erreur physique dans la machinerie matérielle. Ce type d’erreur doit être écarté dans le jeu de l’imitation puisque tout ce qui a trait aux machines réelles a été éliminé d’office. Turing précise [244]:

> **«Dans les discussions philosophiques, on préfère ignorer la possibilité d’erreurs semblables et l’on discute donc de “machines abstraites”. Ces machines abstraites sont des fictions mathématiques plutôt que des objets physiques. Par définition elles sont incapables d’erreurs de fonctionnement. En ce sens, on peut vraiment dire que “les machines sont incapables de faire des erreurs”».**

L’autre type d’erreur est l’erreur de conclusion. Celle-ci n’est pas le fait de la machine mais de celui qui *interprète* les résultats de la machine. C’est ce que fait remarquer Turing quand il dit [245]:

> **«Les erreurs de conclusion ne peuvent survenir que lorsqu’une signification est attachée aux signaux de sortie de la machine. La machine peut, par exemple, imprimer des équations mathématiques ou des mots en anglais. Quand une proposition fausse est imprimée, nous disons que la machine a commis une erreur de conclusion. Il n’y a clairement pas de raison pour dire qu’une machine ne puisse pas commettre ce type d’erreur. Elle pourrait ne rien faire d’autre que d’imprimer “0 = 1”. Pour prendre un exemple moins pervers, elle pourrait avoir une méthode pour tirer des conclusions par induction scientifique».**

La machine simule l’erreur mais ne fait pas d’erreur, tandis que l’interrogateur croit connaître la vérité et commet de ce fait une erreur. La stratégie de la machine au jeu de l’imitation est donc d’essayer de *favoriser* les erreurs de conclusion.

Turing rapporte directement cette stratégie à la sexualité puisqu’il parle ironiquement à son propos d’attitude “perverse”. Turing ironise en généralisant la signification du terme, qui ne signifie plus ici un mode pathologique de plaisir né d’une appréhension jugée anormale de ce qui fait la différence *entre* les sexes mais un mode de plaisir né de l’appréhension de la différence entre ce qui relève

---

[243] **A. M. Turing**, “Computing Machinery and Intelligence”, p. 449.
[244] **A. M. Turing**, “Computing Machinery and Intelligence”, p. 449.
[245] **A. M. Turing**, “Computing Machinery and Intelligence”, p. 449.

du sexuel - l'humain - et ce qui n'en relève pas - le mécanique -. Favoriser les erreurs de conclusion est donc bien de l'ordre de la perversion puisqu'une telle stratégie tend à occulter la différence entre le sexuel humain et l'asexuel mécanique.

Mais Turing fait remarquer qu'une méthode basée sur l'induction est seulement *moins* perverse. Où se situe la "perversion" dans cet exemple ? Comme nous l'avons vu, la machine, en simulant l'erreur, retrouve ce qui faisait la stratégie de la femme et qui faisait son "honnêteté". Ce retour consiste à transformer le contenu de la stratégie de la femme en une simple forme : il s'agit de faire croire que l'on *manifeste* son identité dans les réponses et non pas que l'on essaye de la *cacher*. La "perversité" n'est donc pas l'absence "d'honnêteté" : la perversité consiste à retrouver l'honnêteté par des voies détournées.

En gardant ainsi la simple forme de la stratégie de la femme, la référence à la différence des sexes n'est pas abolie mais seulement *déplacée*. C'est ce qu'il faut essayer de comprendre maintenant.

133. 3. Caractère féminin de la simulation de l'erreur

On sait que l'absence de stratégie de la femme conduit nécessairement à son élimination rapide. C'est pourquoi, après la première allusion à la femme qui établit les règles du jeu de l'imitation, Turing, dans la deuxième et dernière allusion à la femme de l'article, élimine tout ce qui pourrait ressembler à une pensée de la part de la femme, en répondant à ce qu'il appelle "L'objection théologique". Celle-ci prend une forme étrange et très certainement ironique, bien que cette ironie soit fort significative quand on étudie le texte dans son mouvement. Voici comment débute l'objection théologique[246] :

> **«La pensée est une fonction de l'âme immortelle de l'homme. Dieu a donné une âme immortelle à chaque homme et à chaque femme, et non pas à un quelconque animal ou à des machines. Donc aucun animal ni aucune machine ne peuvent penser. Je suis incapable d'admettre une partie quelconque de ceci, mais je vais tenter de répondre en termes théologiques».**

Turing fait remarquer alors combien le dogme chrétien est arbitraire et

---

[246] **A. M. Turing**, "Computing Machinery and Intelligence" , p. 443.

invoque ici, de façon ironique, un autre dogme, qu'il prête à la religion musulmane [247]:

> **«Comment les chrétiens considèrent-ils le point de vue musulman, selon lequel les femmes n'ont pas d'âme ?».**

Même si cette remarque doit être prise *cum grano salis*, il faut cependant noter que c'est précisément au sujet de l'âme de la femme que Turing tourne l'objection théologique en dérision. La fonction de l'argument théologique est cependant claire : il permet de renvoyer la femme à son corps, c'est-à-dire à ce qui relève du physique. On vient de voir cependant que la stratégie de la femme n'était pas abolie mais seulement déplacée. Aussi peut-on dire qu'il y a dans l'attitude de la machine comme la présence d'un corps réduit à une pure forme. C'est ce corps réduit à une simple forme vide qui permet de retrouver le point de vue "honnête" de la femme et de se placer du point de vue du genre humain. Il faudra essayer de comprendre le rôle de ce corps fantôme autour duquel tourne les questions d'honnêteté et de perversité.

Pour le moment, contentons-nous de remarquer que la même stratégie d'honnêteté conduit, dans le cas de la femme, à lui enlever toute âme et, dans le cas de la machine, à lui en donner une. C'est cette conclusion apparemment paradoxale qu'il faut essayer de justifier.

134. Le changement de stratégie propre à la machine et la nature de la psychologie

Remarquons que c'est le changement de stratégie propre à la machine qui permet de quitter le domaine de la différence des sexes et d'atteindre le but du jeu de l'imitation. Il y a donc deux stratégies pour la machine. Dans le cas arithmétique, la machine imite le *contenu* de la stratégie de la femme, son "honnêteté" et commet ainsi des erreurs qui sont interprétées comme relevant du comportement des êtres humains en général. Dans le cas d'une question non-arithmétique, la machine imite la *forme* de la stratégie de l'homme, sa capacité

---

[247] **A. M. Turing**, "Computing Machinery and Intelligence" , p. 443.

d'imitation et répond comme si elle était une femme. La machine imite donc soit la stratégie de la femme soit le comportement de la femme selon que l'on se place du point de vue du contenu ou de la forme. Pourtant la machine ne ressemble physiquement en rien à la femme : l'observateur extérieur en conclut que les différences physiques sont négligeables. Il reste donc à montrer en quel sens on peut considérer, à partir d'un point de vue extérieur au jeu, les différences physiques existant entre les joueurs comme négligeables.

2. La nature physique des joueurs

Turing, quand il décrit les machines susceptibles de jouer au jeu de l'imitation, tente de montrer en quel sens il lui paraît légitime de restreindre le domaine envisagé aux seuls ordinateurs digitaux. Il faut, en vue de saisir son argument, revenir un instant sur le concept de machine pour voir si la restriction qu'il opère est légitime. Turing oppose, quand il décrit les machines susceptibles de participer au jeu de l'imitation, les machines discrètes aux machines continues. C'est par rapport à la notion de machine de Turing que cette distinction est envisagée.

Il nous faut donc, pour étudier la nature physique des joueurs, exécuter, comme Turing, une sorte de va-et-vient entre l'extérieur et l'intérieur du jeu.

21. Les rapports entre machines discrètes

Turing commence par envisager le rapport d'imitation entre machines discrètes.

21. 1. Le jeu n°3 entre machines discrètes

On sait que les machines universelles ont la capacité d'imiter n'importe quelle machine. C'est le même trait caractéristique que Turing accorde aux ordinateurs digitaux, quand il fait remarquer que l'ordinateur digital peut imiter n'importe quel type de machine discrète. L'ordinateur digital ressemble donc, par construction, à la notion de machine universelle, sans en avoir le caractère infini.

Turing définit de façon générale les machines discrètes en faisant remarquer que leur mouvement s'opère par saut successif, permettant de définir des états isolés les uns par rapport aux autres[248] :

> **«Ce sont des machines qui se meuvent par bonds soudains ou qui sautent d'un état bien défini à un autre. Ces états sont suffisamment distincts pour que soit éliminée la possibilité d'une confusion entre eux».**

Aussi le concept d'imitation universelle peut-il ici servir à caractériser la relation qui unit l'ordinateur digital à toutes les machines discrètes. C'est pourquoi Turing mentionne une autre formulation du jeu , que nous appellerons jeu n°3, et dont les règles sont les suivantes : une machine discrète quelconque prend la place de la femme en B, tandis que l'ordinateur digital prend la place de l'homme en A. L'interrogateur, dit Turing, ne peut pas faire la différence entre les deux joueurs [249]:

> **«Le jeu de l'imitation pourrait alors être joué avec la machine en question à la place de B, l'ordinateur digital qui l'imite en A et l'interrogateur serait incapable de les distinguer».**

En effet, dans le cas où l'on arrive à programmer l'ordinateur pour imiter une machine discrète donnée, il n'y a plus de différence entre eux puisque l'ordinateur est devenu, du point de vue formel, la réalisation même de cette machine discrète. La relation, de nature mathématique, qu'entretient l'ordinateur digital et les machines discrètes en général peut-elle se généraliser à d'autres types de machines ?

21. 2. Le jeu n°4 entre l'ordinateur et l'homme

Turing commence par mentionner le fait que l'ordinateur digital peut simuler le calcul effectué par un être humain [250]:

> **«Le lecteur doit accepter comme un fait que des ordinateurs digitaux peuvent être construits et ont assurément été construits selon les principes que nous avons décrits et qu'il peuvent en fait simuler de très près les actions d'un calculateur humain».**

---

[248] **A. M. Turing**, "Computing Machinery and Intelligence", p. 439.
[249] **A. M. Turing**, "Computing Machinery and Intelligence", p. 441.
[250] **A. M. Turing**, "Computing Machinery and Intelligence", p. 438.

A partir de là, Turing mentionne l'existence d'un nouveau jeu, que nous appellerons le "jeu n°4" [251]:

> **«Fixons notre attention sur un ordinateur particulier C. Est-il vrai qu'en modifiant cet ordinateur pour qu'il ait une mémoire adéquate, en accroissant sa vitesse d'exécution et en lui fournissant un programme approprié, on peut faire jouer à C le rôle de A dans le jeu de l'imitation, le rôle de B étant joué par un homme ?».**

Le jeu n°4 a donc le même but que la thèse de Turing n°2 qui identifiait l'être humain et la machine universelle, sauf qu'ici, l'identification s'opère au sein de la réalité physique entre l'être humain et la plus fidèle copie de la machine universelle au sein de l'univers matériel, l'ordinateur digital.

Turing remarque immédiatement que l'analogie qu'il vient de construire entre l'être humain et l'ordinateur n'est qu'une "fiction commode". En quel sens s'agit-il d'une fiction ?

22. Le jeu n° 5 et la réalité continue

Turing considère que tout ce qui fait partie de la réalité physique est de nature continue, y compris les ordinateurs digitaux [252]:

> **«Les ordinateurs digitaux pris en considération dans la dernière section peuvent être rangés dans la classe des "machines à états discrets". Ce sont des machines qui se meuvent par bonds soudains ou qui sautent d'un état bien défini à un autre. Ces états sont suffisamment distincts pour que soit éliminée la possibilité d'une confusion entre eux. Au sens strict, il n'existe pas de machines semblables. Tout se meut continûment. Mais il y a beaucoup d'espèces de machines que l'on peut *considérer* comme des machines à états discrets».**

En tant qu'entités physiques, les machines sont *toutes* analogiques, c'est-à-dire qu'elles reposent toutes sur un mode de fonctionnement continu, impliquant la manipulation de quantités physiques . Mais il est possible de distinguer parmi elles deux types : celles que l'on peut considérer comme discrètes et les autres. Turing dans "Computing Machinery and Intelligence" donne un exemple de ces dernières, l'analyseur différentiel[253].

---

[251] **A. M. Turing**, "Computing Machinery and Intelligence", p. 442.
[252] **A. M. Turing**, "Computing Machinery and Intelligence", p. 439.
[253] **A. M. Turing**, "Computing Machinery and Intelligence", pp. 451-452.

L'analyseur différentiel est une machine qui calcule de façon analogique des équations différentielles. C'est Lord Kelvin qui l'inventa, à partir de la machine qu'il avait inventé pour prévoir les marées et que nous avions déjà mentionné. L'intérêt d'une telle machine analogique en comparaison des machines discrètes existant à l'époque comme la machine digitale de Babbage dont Turing parle dans son article[254], est sa rapidité[255]. Vannevar Bush, un ingénieur américain, redécouvrit dans l'entre-deux guerres l'idée de Lord Kelvin et l'améliora : plusieurs analyseurs différentiels furent construits aux États-Unis et le département de mathématiques de Cambridge en Angleterre en acheta un. Turing a donc pu voir fonctionner et utiliser cette machine. Un autre analyseur différentiel avait été construit peu auparavant à l'Université de Manchester par le physicien Douglas Hartree[256]. L'analyseur différentiel eut une grande importance durant la guerre, parce que c'est avec lui que l'on calculait les tables balistiques nécessaires à l'artillerie. Il ne fut supplanté aux États-Unis qu'à la fin de 1944, à partir de l'apparition de machines à calculer électroniques[257].

Comment réussir à isoler au sein des machines qui, en tant qu'entités physiques, sont *toutes* analogiques, un type de machine qui soit considéré comme digital ?

D'un point de vue intellectuel, on considère qu'une machine est digitale quand ses opérations de base sont les opérations fondamentales de l'arithmétique : addition, soustraction, multiplication et division. De ce point de vue, l'analyseur différentiel n'est pas une machine que l'on puisse considérer comme digitale parce que ses opérations fondamentales sont la différentiation, l'intégration et la résolution de certaines équations différentielles. Aussi une opération aussi

---

[254] **A. M. Turing**, "Computing Machinery and Intelligence", p. 439.

[255] Cf **H. Goldstine**, *The Computer from Pascal to Von Neumann*, Princeton University Press, Princeton, 1972, p. 142-143. Turing lui-même insiste sur le fait que la machine de Babbage est plus rapide que'un calculateur humain mais beaucoup moins rapide (cent fois moins) qu'un ordinateur difgital. Cf **A. M. Turing**, "Computing Machinery and Intelligence", p. 439.

[256] Cf. **H. Goldstine**, *The Computer from Pascal to Von Neumann*, op. cit., p. 95. Notons que Turing cite, dans "Computing Machinery and Intelligence" p. 450, le livre de Hartree *Calculating Instruments and Machines* et se déclare en complet accord avec l'auteur quand celui-ci évoque la possibilité d'une machine intelligente, possibilité qui n'était pas apparue clairement à Babbage.

[257] Cf **H. Goldstine**, *The Computer from Pascal to Von Neumann*, op. cit., pp. 165-166.

“simple” que la multiplication exige, dans le cas de l’analyseur différentiel, d’être calculé par le biais de l’intégration. Bref, pour une machine digitale, toutes les opérations autres que les opérations fondamentales de l’arithmétique sont considérées comme transcendantes, tandis que c’est l’inverse pour les machines analogiques[258]. C’est donc la différence entre l’infini du discret propre à l’arithmétique finitaire et l’infini du continu propre à l’arithmétique infinitaire qui permet de fonder la différence entre les deux types de machines.

Mais, du point de vue physique, les deux types de machines sont, pour Turing, *indiscernables*. Il propose pour l’établir une cinquième version du jeu de l’imitation dans laquelle l’interrogateur doit essayer de faire la distinction entre l’analyseur différentiel d’une part et l’ordinateur digital d’autre part, en supposant que les deux machines “répondent” à une question portant sur le calcul des décimales de π [259]:

> **«Il ne serait pas possible à un ordinateur digital de prévoir exactement quelles réponses l’analyseur différentiel pourrait donner à un problème mais il serait tout à fait capable de donner la bonne sorte de réponse. Par exemple, si on lui demandait la valeur de π (à peu près 3,1416), il serait raisonnable de choisir au hasard entre les valeurs 3,12; 3,13; 3,14; 3,15; 3,16 avec, disons, des probabilités de 0,05; 0,15; 0,55; 0,19; 0,06. Dans ces conditions, il serait très difficile à l’interrogateur de faire la différence entre l’analyseur différentiel et l’ordinateur digital.»**

Quelles sont les places respectives des participants à ce jeu par rapport au jeu n°1?

L’interrogateur, selon sa place et sa fonction habituelles, est censé ne pas pouvoir opérer la distinction qui lui est proposée.

L’ordinateur digital occupe la place de l’homme (A) qui imite, dans la mesure de ses capacités, la femme (B). L’analyseur différentiel, comme la femme, n’imite que lui-même. Aussi peut-on ranger l’analyseur différentiel du point de vue de la femme et de son “l’honnêteté” stratégique.

On peut donc construire le tableau suivant, représentant le jeu de l’imitation dans ses cinq versions successives :

---

[258] Cf. **H. Goldstine**, *The Computer from Pascal to Von Neumann*, op. cit., p. 142.
[259] **A. M. Turing**, “Computing Machinery and Intelligence”, pp. 450-451.

	A	B
Jeu n° 1	Homme	Femme
Jeu n°2	Ordinateur Digital	Femme
Jeu n°3	Ordinateur Digital	Machine Discrète
Jeu n°4	Ordinateur Digital	Homme
Jeu n°5	Ordinateur Digital	Analyseur Différentiel

Au vu du tableau, on s'aperçoit, en prenant l'ordre chronologique d'apparition des jeux, que l'ordinateur digital, à partir du jeu n°2, a finalement été comparé avec toutes les autres entités, humaines ou non-humaines, discrètes ou continues.

Quelle conséquence tirer de cette analyse du point de vue de la "psychologie" que l'on peut prêter à l'ordinateur dans le jeu de l'imitation et ses métamorphoses successives?

La seule façon de parvenir à montrer que du point de vue intellectuel comme du point de vue matériel, l'ordinateur digital ne peut pas être distingué d'un être humain exige que l'on prenne *en même temps* en considération le point de vue intérieur et extérieur au jeu. Ce point de vue doit donc se situer *par delà la distinction de l'intellectuel et du physique*, puisque le rapport établi entre l'être humain et l'ordinateur digital du point de vue intellectuel et discret se fonde sur une identité commune du point de vue de leur nature physique et continue.

Comment rendre compte de ce point de vue dans le cadre du jeu de l'imitation ?

23. Par-delà la distinction du physique et de l'intellectuel

On avait remarqué que le jeu de l'imitation ne pouvait avoir de sens que si l'on distinguait entre ce qui se passe dans le jeu et ce qui se passe hors de lui et que Turing jouait de cette différence pour se situer dans une position de "survol" qui lui permettait d'occuper les deux points de vue sur le jeu. Aussi pouvait-il *à la fois* faire la différence entre la nature physique des joueurs et considérer que cette différence ne pouvait pas apparaître dans le jeu. C'est en effet cette position tout à fait particulière qui lui permet de répondre "non" à la question de savoir si l'interrogateur pourra ou non deviner s'il a affaire à un être humain ou à un ordinateur digital dans une partie de jeu n°2. Comment réussir à justifier cette position de survol ?

On a vu que l'argumentation tourne autour de l'absence de stratégie véritable accordée à la femme et de ce que les deux stratégies de la machine imitent dans leur forme et leur contenu l'attitude de la femme. On va voir que les raisons qui ont permis à Turing d'adopter ce présupposé touchant l'attitude de la femme doivent être reliées à l'invention du concept de machine de Turing : la capacité féminine d'engendrement biologique serait analogue pour Turing à la capacité masculine d'invention. Il faut donc tenter d'éclaircir maintenant les conditions qui ont rendu possible l'invention du concept de machine universelle.

Chapitre V

Interprétation psychologique du jeu de l’imitation

On a vu que la machine était susceptible de mener deux stratégies au jeu de l’imitation, l’une féminine et l’autre masculine et que c’était en cela que résidait sa “perversité”[260]. Cette “perversité” peut cependant s’interpréter de deux façons différentes. On peut en effet considérer soit que la machine se situe au-delà de la différence des sexes soit qu’elle réintègre par un biais détourné le cadre de la différence des sexes, biais que Turing appelle “pervers”.

On remarque d’autre part qu’il y a un rapport entre la “perversité” de la machine et le point de vue de l’observateur extérieur au jeu, par exemple le lecteur, qui, lui aussi, doit être capable à la lecture de l’article de dépasser la différence des sexes et de constituer l’intelligence en un concept véritable. Deux conséquences en découlent. Du point de vue de la machine, le comportement qu’elle manifeste la rend identique au point de vue de l’observateur extérieur et c’est pourquoi on peut lui attribuer une intelligence. Du point de vue de l’observateur extérieur, on peut dire que le lecteur de l’article adopte un point de vue “pervers” puisque lui aussi se situe par-delà la différence des sexes. Cependant, avant le lecteur de l’article, c’est *Turing lui-même* qui a dû adopter ce point de vue “pervers” puisque c’est lui qui a inventé le jeu de l’imitation. La “perversité” aurait ainsi à voir avec l’invention du concept de machine de Turing lui-même. Aussi peut-on voir dans le jeu de l’imitation *le récit autobiographique*

---

[260] Cf. **A. M. Turing**, “Computing Machinery and Intelligence”, op. cit., p. 449 : «Pour prendre un exemple moins pervers, elle [la machine] pourrait avoir une méthode pour tirer des conclusions par le biais de l’induction scientifique».

de la façon dont Turing a inventé le concept de machine de Turing en s'identifiant, comme le lecteur de l'article est invité à le faire après lui, à une machine, ce qui implique d'adopter une attitude "perverse" à l'égard de la différence des sexes.

Un indice va dans le sens de cette interprétation psychologique. Turing attribue une "perversité" dans le cas, apparemment "honnête", d'une machine capable de tirer des conclusions par induction scientifique. Or une machine capable de tirer des conclusions par induction scientifique est une machine capable d'apprentissage. Si l'interprétation autobiographique est recevable, on peut voir dans le jeu de l'imitation les étapes de l'apprentissage qui furent nécessaires à Turing pour inventer le concept de machine, étapes dont le franchissement aurait été rendu possible par cette attitude perverse. Il est nécessaire pour parvenir à cette conclusion, de revenir à l'identification de l'observateur extérieur et de la machine, c'est-à-dire de revenir à *l'invention* du concept de machine de Turing.

Si Turing parlait explicitement des voies d'accès qui l'ont conduit au concept de machine de Turing, on pourrait se forger une meilleure idée des conditions qui en ont rendu possible l'invention. Mais Turing n'en dit rien explicitement. La seule indication qu'il ait, de son vivant, confié à R. Gandy [261] concernant son invention est la suivante : ce fut allongé sur une pelouse, pendant une halte après une longue course à pied[262] qui l'avait amené de Cambridge à Grantchester, que lui vint à l'esprit «l'idée principale» de "On Computable Numbers …", un jour de l'été 1935. A l'évidence, cette remarque biographique ne suffit pas à expliquer l'invention du concept.

On remarque cependant qu'une autre voie d'accès demeure possible : il est

---

[261] Cf. **R. O. Gandy**, "The Confluence of Ideas in 1936" dans [*The Universal Turing machine*, **R. Herken ed.**, Oxford Science Publications, Oxford University Press, 1988] p. 82.

[262] Comme le remarque le biographe de Turing, la course à pied avait une connotation sexuelle pour Turing qui, dans son adolescence y voyait un substitut à la masturbation. Plus tard, il avait inventé un jeu qui combinait la course et le jeu d'échecs : entre chaque coup, il fallait courir jusqu'à ce que l'adversaire joue à son tour, la course empêchant toute concentration entre les coups. Turing fut même pressenti pour faire partie de l'équipe britannique de course de Marathon pour les Jeux Olympiques qui se déroulèrent en 1948 en Grande-Bretagne. Il y renonça en raison d'une blessure à la hanche. Cf **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., pp. 57, 96, 346 et 387.

question à plusieurs reprises dans “Computing Machinery and Intelligence” de l’accès à l’universalité, *précisément sous l’aspect de l’induction scientifique*. C’est donc cette question qui est commune à *l’invention* du concept de machine de Turing et à ce que Turing suppose de *l’apprentissage* “pervers” des machines qu’il nous faut maintenant étudier parce que c’est elle qui peut nous donner des indications sur la façon dont le concept a été inventé.

1. L’invention du concept de machine de Turing

11. L’induction scientifique dans “Computing Machinery and Intelligence”

Il y a trois allusions au principe de l’induction scientifique dans l’article de Turing.

La première se situe dès les premières lignes du texte, quand il fait remarquer qu’il est absurde de vouloir répondre à la question : “Les machines peuvent-elles penser?” par voie de sondage. L’usage n’est pas le meilleur guide pour essayer de répondre à la question, car il ne permet pas la précision requise pour la détermination d’un concept. On sait d’autre part que cette question suppose l’universalité de la distinction, établie par induction, entre le physique et l’intellectuel. Or cette universalité est trompeuse, puisque la question est mal posée. C’est pourquoi il faut abandonner l’idée d’un sondage et tenter de construire une véritable expérience, celle du jeu de l’imitation. L’induction contient donc des pièges et c’est la question du rapport entre l’induction et l’universalité qu’il faut réussir à éclaircir.

La deuxième allusion[263] au principe de l’induction porte précisément sur le rapport entre l’induction et l’universalité des conclusions qui peuvent en découler. Turing fait remarquer que l’on dénie aux ordinateurs toutes les caractéristiques que l’on accorde aux êtres humains (gentillesse, beauté, possibilité de se lier d’amitié ou de tomber amoureux, etc) parce que, en appliquant le principe d’induction, on conclut que les machines sont habituellement extrêmement limitées dans leur fonction. Ce qui leur manque le

[263] Elle se situe dans le paragraphe qui s’intitule “*Arguments from various disabilities*”.

plus est précisément de pouvoir s'adapter à différentes tâches, comme le font les êtres humains. L'induction consiste alors à conclure qu'il est nécessaire pour toute machine d'être limitée dans ses fonctions et dans ses buts. C'est une application *erronée* du principe d'induction, parce qu'il est possible de concevoir une machine *universelle*, capable de s'adapter à tous les buts arithmétiques, comme Turing l'a montré dans son article de 1936. Pour Turing, qui a justement réussi à exhiber l'universalité du concept de calculabilité par le biais de la thèse de Turing, il est possible de conclure légitimement que cette induction est erronée parce qu'elle repose sur un nombre d'exemples de machines trop restreint.

Immédiatement après avoir énoncé une conclusion similaire au sujet de l'universalité des ordinateurs digitaux, Turing revient sur le principe d'induction et en propose deux exemples. Ces deux exemples mettent en scène des enfants qui appliquent, de façon inconsciente dit Turing[264], le principe en question.

Le premier exemple est le suivant. Quand un enfant qui s'est brûlé a ensuite peur du feu et manifeste cette peur en évitant celui-ci, il applique à bon escient le principe d'induction. Cette application du principe d'induction est légitime, parce qu'elle porte sur des données naturelles, le rapport entre le corps de l'enfant et le feu. Ce rapport s'exprime dans la *passion* que constitue la peur. La transmission de la proposition générale concernant le feu s'effectue donc par le biais d'un canal émotionnel, la peur. Turing fait remarquer ensuite que, lorsque l'induction porte sur des données d'ordre culturel, l'induction est plus difficile à réaliser : il fait ainsi remarquer que la plupart des petits Anglais considèrent qu'il est idiot d'apprendre le français, puisque tout le monde parle anglais. Dans le cas des données culturelles, il est donc facile de se méprendre sur l'universalité de la proposition dégagée par induction quand on prend pour contenant universel ce qui n'est que contenu particulier : c'est le cas pour l'exemple de la langue anglaise qui n'est qu'une langue parmi d'autres et non pas la langue universelle. Il y a cependant possibilité de faire une induction à bon escient, comme Turing le sait lui qui a réussi à mettre au jour l'universalité du concept de calcul. Mais Turing fait remarquer qu'il est difficile d'opérer une bonne induction sur des matériaux

---

[264] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 448.

relevant de la culture [265]:

> **«Les travaux et les coutumes de l'humanité ne semblent pas être un matériau très adapté à l'application de l'induction scientifique. On doit mener une enquête sur une très grande partie de l'espace-temps, si l'on veut obtenir des résultats fiables».**

Le rapport entre ces deux exemples n'est pas immédiat, puisque Turing ne donne pas de critère objectif qui permettrait d'appliquer à bon escient le principe d'induction. Et cela pour une bonne raison : l'accumulation de l'expérience ne peut pas résoudre le problème de l'induction. Considérer que l'universalité d'une proposition a été atteinte après application du principe d'induction revient en effet à présupposer l'universalité de la proposition en question puisqu'il est impossible de passer en revue tous les cas possibles.

On se trouve donc devant la situation suivante : dans le cas de l'enfant qui s'est brûlé, l'induction est légitime mais pas dans le cas de l'enfant qui croit à l'universalité de sa langue maternelle. Par ailleurs, on sait que l'universalité culturelle existe, puisque l'universalité du concept de machine de Turing est avérée. Aussi est-ce la manière dont cette universalité a été acquise qui reste mystérieuse.

12. Le sacrifice de "Casabianca"

En fait, le chaînon manquant entre l'exemple de l'enfant brûlé et celui de l'enfant qui croit à l'universalité de la langue anglaise se situe ailleurs dans l'article de Turing, quand il fait une allusion au poème anglais qui s'intitule "Casabianca"[266].

Il s'agit d'un poème écrit par une femme, Felicia Dorothea Hemans[267], qui exalte l'héroïsme d'un enfant à travers un fait d'arme français[268] : on rapporte que pendant la campagne d'Égypte de 1798, lors de la bataille du Nil en rade d'Aboukir où les Français furent vaincus par la flotte anglaise commandée par

---

[265] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 448.
[266] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 457.
[267] Née en 1793 et morte en 1835.
[268] C'est ce qui explique sans doute que le second exemple de Turing porte sur la différence entre l'anglais et le français.

l'amiral Nelson, la manœuvre du côté français était assurée par le capitaine Louis Casabianca et que, même après la mort de l'amiral Brueys d'Aigaïlliers, il continua à se battre jusqu'au bout. Son fils âgé de treize ans, Giacomo, refusa de quitter le navire et mourut avec son père. Le poème décrit la discipline dont fait preuve Giacomo qui appelle trois fois son père pour lui demander de quitter son poste, sans savoir que ce dernier est déjà mort et qu'il ne peut pas lui répondre; les flammes finissent par atteindre les réserves de poudre et dans l'explosion du navire, les fragments du corps de Giacomo sont éparpillés sur la mer. La strophe la plus souvent citée est la première [269]:

> **«Le garçon se tint debout sur le pont en feu**
> **duquel tous sauf lui avaient fui;**
> **Les flammes éclairant les débris de la bataille**
> **brillaient autour de lui au-dessus des morts»**
>
> **[*"The boy stood at the burning deck***
> ***Whence all but he had fled;***
> ***The flame that lit the battle's wreck***
> ***Shone round him o'er the dead"*]**

Le chaînon manquant que constitue ce poème permet de mettre en rapport les deux exemples donnés par Turing et d'éclairer la difficile question du rapport de l'induction et de l'universalité.

Dans le cas de "Casabianca", Giacomo aurait dû fuir le navire, comme tous les autres marins et soldats encore vivants, appliquant en cela à bon escient le principe d'induction qui consiste, comme le dit Turing dans le premier exemple, à avoir peur du feu[270]. Pourquoi Giacomo refuse-t-il d'appliquer le principe d'induction ? Par fidélité à l'égard de son père dont il ignore qu'il est mort. Dès lors, cet exemple en devient plus qu'un exemple particulier : c'est par le sacrifice et la mort qu'il prend une valeur universelle. Le mystère du rapport entre l'induction et l'universalité passe donc par ce biais : l'expérience portant «sur une grande partie de l'espace-temps» dont parlait Turing est ainsi une expérience

---

[269] La strophe citée est si connue en Angleterre qu'elle est reproduite dans *The Oxford Dictionnary of Quotations*, Oxford University Press, Oxford, 3rd edition, 1979, p. 244.

[270] Le feu étant pris ici au propre et au figuré, puisque, en anglais comme en français, on dit que le vaisseau de Casabianca est en feu après avoir essuyé le feu des navires anglais.

radicale, celle de la fin de l'espace-temps lui-même, à savoir la mort. L'expérience de ce sacrifice possède deux traits caractéristiques : il s'agit d'un sacrifice par le *feu* qui joue le rôle d'un *message*. Giacomo Casabianca suit l'exemple de son père : son sacrifice a pour but d'assurer la transmission d'un message de fidélité à son père en mettant en péril son corps. Ce sont précisément les deux stratégies de la machine "perverse" qui se trouvent représentées dans cet exemple de sacrifice : stratégie féminine d'abord qui consiste à tenter de sauver son corps en espérant recevoir l'ordre de quitter le bateau en perdition, stratégie masculine ensuite qui consiste à le sacrifier pour en préserver la pure forme, le contenant. Dans l'autre exemple au contraire, les petits Anglais, considèrent leur langue maternelle comme universelle parce que leur apprentissage de la langue s'est opéré par un canal de nature émotionnel. Ils ne disposent pas d'un contenant qui leur permettrait de constituer la langue anglaise en contenu particulier, parce qu'ils n'ont pas fait, comme dans l'exemple du poème, l'épreuve du sacrifice. Un *contenu sans contenant*, voilà donc ce qui rend le principe d'induction illégitime. Dans le cas du sacrifice de Giacomo Casabianca au contraire, le contenant prime sur le contenu : le sacrifice consiste à faire taire en soi le canal émotionnel de la passion. Cette épreuve du sacrifice se confond avec la discipline absolue de Giacomo qui reste à son poste tant qu'il n'a pas reçu l'ordre de quitter le navire[271]. Dans le contexte de la théorie du calcul, c'est cette discipline absolue qui permet d'atteindre l'universalité, c'est-à-dire, dans le cas de Turing, de parvenir à l'invention du concept de machine universelle. C'est ce qu'explique Turing dans un texte juste antérieur à "Computing Machinery and Intelligence" [272]:

> **«Transformer un cerveau ou une machine en une machine universelle est la forme la plus extrême de la discipline».**

---

[271] Remarquons aussi qu'un navire de guerre peut en anglais se dire [*man of war*] mais que les bateaux sont néanmoins du genre *féminin*. Il y a donc ici, entre l'aspect sémantique et l'aspect syntaxique de l'expression, un étrange mélange des genres qui a pu susciter le souvenir du poème chez Turing.

[272] **A. M. Turing**, "Intelligent Machinery", op. cit., p. 49. On avait noté, au cours de la première partie, que le concept de machine universelle avait une intérêt du point de vue psychologique avant d'en avoir un d'un point de vue mathématique : on en voit dans cette citation la confirmation.

Plus précisément, si l'on revient à l'exemple de Giacomo Casabianca, le sacrifice consiste en ce que la *peau* soit brûlée, c'est-à-dire ce qui permet de faire une distinction entre l'intérieur et l'extérieur du corps. Ainsi l'induction est-elle légitime[273], si on prend le risque de sacrifier ce qui fait la différence entre l'intérieur et l'extérieur du corps : la peau. C'est donc par rapport à la peau qu'il est possible de rendre compte du passage, par le biais de ce qui fait exemple, des cas particuliers naturels à l'universalité culturelle, et tout particulièrement à l'universalité de la langue, comme le note le deuxième exemple évoqué par Turing, celui des petits Anglais monolingues.

On ne peut pas ne pas se demander quelle fut, pour Turing, l'expérience personnelle de type sacrificiel qui lui permit de dégager l'universalité du concept de machine de Turing, rendant ainsi possible, dans le domaine culturel, une induction jusqu'à l'universel. L'ordinateur digital représente en effet comme le rêve d'une langue universelle, réalisé après une "grande enquête dans l'espace-temps". Quelle a été l'enquête personnelle de Turing dans "l'espace-temps" ?

Que l'on se rappelle les étapes successives de notre argumentation : la position de survol propre à l'observateur extérieur (et à Turing) lui permet de faire une différence entre l'intérieur et l'extérieur du jeu. Cette position de survol ne peut se justifier sans faire appel à l'induction scientifique et à la différence qu'elle rend possible entre un contenu et un contenant. Turing rapporte cette différence dans l'exemple de "Casabianca", à un sacrifice qui porte sur la peau et la différence qu'elle permet d'instaurer entre l'intérieur et l'extérieur du corps. On peut conclure que la peau entretient avec la position de survol qui oppose un intérieur à un extérieur, un lien secret. C'est ce lien qu'il faut réussir à mettre au jour.

2. La peau comme interface entre le physique et l'intellectuel

Contrairement à ce que l'on pourrait supposer au premier abord, il existe

---

[273] Dans le cas du poème, l'induction "sacrificielle" consiste à devenir totalement abstrait, c'est-à-dire privé de corps.

toute une méditation sur le statut de la peau dans l'article de Turing. C'est sur elle qu'il faut maintenant fixer l'attention.

Les allusions à la peau, dans "Computing Machinery and Intelligence" sont à la fois nombreuses et peu visibles, dans la mesure où elles ont une fonction seulement pédagogique qui a pour but de clarifier le propos en fournissant des exemples. Elles sont au nombre de sept, dispersées au fil des vingt-neuf pages de l'article[274], sans d'ailleurs que soit toujours expressément mentionné le terme de peau. Essayons de dégager les fonctions de ces allusions.

21. La peau comme manifestation de la nature physique de l'être humain

Il est frappant de constater qu'une fois exposé les règles du jeu de l'imitation, Turing commence dès le départ, par faire une allusion directe à la peau [275]:

> **«Le nouveau problème a l'avantage de tracer une ligne assez nette entre les capacités physiques et intellectuelles d'un homme. Aucun ingénieur ni aucun chimiste ne prétend être capable de produire un matériau indiscernable de la peau humaine. Il est possible que ceci soit réalisé dans l'avenir mais cette invention serait-elle disponible que nous devrions sentir combien rendre plus humaine une "machine pensante" en l'habillant de cette chair artificielle serait hors de propos».**

Pour Turing, ce qui relève du physique dans l'homme se manifeste donc d'abord par l'existence de la peau. C'est pourquoi il place cette notion hors-jeu, puisqu'elle relève précisément de ce qu'il faut éliminer, le simple habillage externe de l'intelligence humaine. C'est en cherchant à éliminer toute référence à la peau que Turing compte montrer que la notion d'intelligence n'est pas le propre de l'homme et qu'elle peut s'incarner dans des substrats physiques très divers, comme les tissus vivants dans le cas de l'espèce humaine ou les composants électroniques dans le cas de l'ordinateur digital.

On doit donc préciser ainsi le but des règles du jeu de l'imitation : si tout contact - et au premier chef le contact tactile - est prohibé entre les joueurs, c'est

---

[274] Voici où elles se trouvent : (1) page 434, ligne 27; (2) p. 436, l. 3; (3) p. 448, l. 8; (4) p. 453, l. 11; (5) p. 454, l.37; (6) p. 456, l. 38; (7) p. 457, l. 17. A celles-ci, il faudrait rajouter l'une des réponses du dialogue proposé par Turing en exemple p. 434 : quand l'homme répond «mes cheveux sont coupés à la garçonne», le terme employé par Turing pour désigner cette coupe est [*shingled*] qui se rapproche fortement du mot [*shingles*] qui désigne une maladie de peau, le zona.

[275] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 434.

bien pour tracer une ligne de démarcation aussi nette que possible entre le physique et l'intellectuel, non pas seulement parce que la peau est la manifestation la plus immédiate de l'existence *physique* de l'être humain mais parce qu'elle joue le rôle plus subtil d'une interface entre le domaine du *physique* et celui de *l'intellectuel*. On ne peut bien comprendre le rôle d'interface que joue la peau sans préciser la place qu'elle occupe dans l'élaboration de l'identité sexuelle chez les êtres humains : c'est d'ailleurs ce à quoi s'emploie Turing, sans doute sans le savoir, dès les paragraphes suivants.

22. La peau comme manifestation de la différence sexuelle

Turing essaye en effet dans les paragraphes suivants, c'est-à-dire dès la deuxième page de l'article, de préciser ce qu'il entend par machine quand il est question du jeu de l'imitation. Il est remarquable de constater qu'au lieu de dévoiler immédiatement sa pensée et de montrer qu'il lui semble légitime de pouvoir réduire la notion de machine à celle d'ordinateur digital, Turing cherche tout d'abord à exprimer de façon très générale les caractéristiques qui permettent de déterminer quel type de machines on peut admettre au jeu de l'imitation.

Turing énonce trois caractéristiques : tout d'abord, n'importe quel type d'ingénierie sera considéré comme exploitable pour la construction de machines; ensuite, même si l'ingénieur ou l'équipe d'ingénieurs qui construisent une machine ignorent son fonctionnement interne parce que sa construction est largement empirique, on considérera qu'il s'agit tout de même d'une machine; enfin, les êtres humains nés de façon naturelle ne seront pas assimilés à des machines.

La troisième caractéristique qui porte sur des questions ayant trait aux modalités humaines de la naissance est de nature négative. Turing fait remarquer qu'il est difficile d'harmoniser ces trois conditions et se fait à lui-même cette curieuse objection [276]:

> **«On pourrait par exemple insister sur le fait que l'équipe d'ingénieurs devrait être toute du même sexe, mais ce ne serait pas vraiment satisfaisant, car il est probablement**

[276] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., pp. 435-436.

possible de construire un individu complet à partir d'une seule cellule, disons de la peau d'un homme».

Comment comprendre cette objection ?

221. L'objection de l'équipe d'ingénieurs

La remarque touchant l'équipe d'ingénieurs, dans laquelle se mêle sans doute un brin d'humour, n'en reste pas moins extrêmement étrange et exige une explication, à la fois parce que Turing la considère comme une objection et parce qu'il prend le temps d'y répondre.

Remarquons tout d'abord que cette objection occupe une place tout à fait particulière dans l'article de Turing puisqu'elle ne se situe pas dans la partie qui regroupe les objections mais qu'elle se trouve placée au début de l'article. Il s'agit donc d'une objection mais qui n'en a pas le statut et que Turing discute presque de façon incidente. Pour ce qui est de l'objection proprement dite, on peut se demander en quoi l'existence d'une équipe d'ingénieurs du même sexe peut apparaître comme une condition *sine qua non* à la construction d'une machine. Car enfin, en quoi le sexe des ingénieurs aurait-il une quelconque influence sur la détermination sexuelle de ce qu'ils construisent ?

Il faut supposer que ce que Turing a dans l'esprit est un engendrement au sens biologique du terme : c'est parce qu'il s'agit d'un engendrement biologique que le sexe des ingénieurs détermine la nature sexuelle de ce qui est créé. On peut donc supposer - et c'est sans doute là que se situe l'humour - que l'objection vise à supprimer la possibilité d'une tricherie de la part de l'équipe d'ingénieurs, tricherie qui consisterait à faire passer pour une création artificielle une "machine" qui aurait été en fait obtenue par une fécondation et une gestation naturelle. En formant une équipe d'ingénieurs du même sexe, on supprime du même coup la différence sexuelle et la possibilité de toute fécondation et gestation naturelle : si l'on s'accordait cette possibilité, on présupposerait finalement ce qu'il faut prouver, à savoir que la machine est indiscernable, d'un certain point de vue, d'avec un être humain. L'objection sous-entend donc que la différence sexuelle entre dans le cadre de ce qui appartient au substrat physique : tenter de supprimer la différence sexuelle entre ainsi dans le projet général de supprimer tout ce qui a trait au substrat physique particulier propre à l'espèce humaine. Et pourtant, Turing n'en reste pas là puisqu'il critique l'objection, alors qu'elle va dans le sens même de ce qu'il défend. Pourquoi ?

Si l'équipe d'ingénieurs est susceptible de tricher et de faire passer ce qui est fécondation et gestation naturelle pour une construction artificielle, c'est que la vraie fonction de l'équipe en question doit être de jouer elle-même, sans le secours de la nature sexuelle des membres de l'équipe, le rôle d'un utérus féminin dans lequel s'élaborerait la constitution progressive du foetus. Or, pour Turing, l'équipe d'ingénieurs peut jouer ce rôle, que les deux sexes soient ou non représentés au sein de l'équipe : c'est pourquoi un ingénieur unique ferait aussi bien l'affaire.

Aussi l'intervention de la peau pour répondre à l'objection devient-elle plus claire. La première fonction de la peau est en effet de séparer l'intérieur de l'extérieur du corps : si on peut reconstruire un individu complet à partir d'une seule cellule de peau, alors il n'y a pas besoin de chercher sous la peau, dans les profondeurs de l'utérus, ce qui se trouve déjà à l'extérieur, à la surface du corps. L'objection touchant le sexe de l'équipe d'ingénieurs n'est donc, pour Turing, pas recevable.

222. Les modes de procréation

Mais la question de la nature sexuelle de la peau se trouve du même coup posée. En effet, la réponse de Turing, comme l'objection qu'elle critique, supprime la différence sexuelle, puisqu'il est possible de reconstruire un individu à partir d'une cellule de sa peau sans fécondation bisexuée. Mais cette reconstruction se situe-t-elle hors du domaine du sexuel ? Rien ne l'indique. Au contraire, il semble que Turing élabore une analogie entre le sexe féminin et l'intérieur du corps d'une part et le sexe masculin et la surface du corps de l'autre : la peau en tant que surface serait au fondement de l'identité sexuelle de l'homme. L'individu construit à partir d'une cellule unique de peau serait alors un *homme*.

Un autre indice présent dans le texte va d'ailleurs dans le même sens.

Comme on l'a déjà remarqué, le but exprimé ouvertement par Turing est non pas de montrer qu'il n'est pas possible de faire la différence entre l'être humain et la machine au jeu de l'imitation, mais qu'il n'est pas possible de faire une différence entre un *homme* et une machine. L'expression utilisée en anglais à ce sujet ne laisse en effet aucun doute [277]:

> **«Le nouveau problème a l'avantage de tracer une ligne assez nette entre les capacités physiques et intellectuelles d'un homme [*of a man*].»**

Plus généralement, on constate que les allusions à l'homme sont fréquentes au cours de l'article. Mais on remarque l'existence d'une différence entre elles : Turing semble en effet utiliser à la fois l'usage générique du terme d'homme (L'Homme, [*Man*]) et son usage particulier (un homme, [*a man*]). En quelles circonstances utilise-t-il l'une ou l'autre ?

La seule fois où Turing emploie le terme d'Homme au sens générique se trouve dans la réponse à la seconde objection, de nature théologique, que Turing appelle "L'objection de "l'autruche"[278]. L'objection est la suivante : admettre l'existence de machines qui pensent serait si terrible qu'il vaut mieux continuer à croire en la supériorité de l'Homme telle que la religion nous l'enseigne. Partout ailleurs dans l'article, il n'est pas question de l'Homme mais d'un homme particulier, contrairement au terme de femme pour lequel Turing n'introduit pas de différence quant au degré de généralité. Ainsi par exemple dans la réponse à l'objection de l'équipe d'ingénieurs du même sexe que nous venons de citer, Turing précise-t-il bien qu'il suffit d'une seule cellule de la peau d'*un* homme pour construire un individu complet.

Pourquoi Turing introduit-il cette distinction entre un usage générique et un usage particulier du terme d'homme ?

On pourrait croire tout d'abord qu'il s'agit d'une précaution : Turing pourrait avoir eu l'intention de relever l'ambiguïté de l'usage commun de la langue, qui range les deux sexes sous le terme qui sert aussi à en désigner un en particulier, l'anglais, comme le français, ayant perdu la distinction existant en

---

[277] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 434.
[278] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 444.

latin entre [*Homo*], l'Homme pris au sens d'être humain et [*Vir*], l'homme par opposition à la femme. Mais en fait, contrairement aux apparences, l'usage du terme d'homme dans son sens particulier a pour but non pas d'insister sur la différence des sexes mais au contraire d'éliminer définitivement toute référence à la femme : c'est toujours du sexe masculin dont il est question dans l'article[279] et jamais de l'Homme pris au sens générique d'être humain, notion qui inclut, tout en l'occultant il est vrai, la différence sexuelle.

Turing, en répondant à l'objection de l'équipe d'ingénieurs, supprime donc la *différence* sexuelle mais ne supprime pas la *référence* au sexuel.

La construction d'un individu à partir d'une cellule unique de peau entretient donc un rapport curieux avec la comparaison entre procréation naturelle des enfants et construction artificielle des machines : elle a, elle aussi, un statut d'interface puisqu'elle est en position médiane par rapport aux deux autres modes d'engendrement. Ainsi y-a-t-il bien deux types de création possible, selon la place qu'ils occupent par rapport à la notion de peau : la première est rendue possible par l'intérieur de la femme, la deuxième est rendue possible par la peau de l'homme. Les entités créées sont alors soit masculines ou féminines dans le cas de la procréation naturelle, soit masculines dans le cas de la procréation artificielle.

Aussi le texte de Turing est-il construit sur une analogie : du point de vue de la position de survol, l'intérieur du jeu est à l'extérieur du jeu ce que, du point de vue de la peau, l'intérieur du corps est à ce qui en est extérieur. Une question continue cependant de se poser : quel rapport la machine entretient-elle avec la détermination sexuelle ?

23. La peau et la machine

La machine possède-t-elle une détermination sexuelle ? C'est par rapport au statut de la peau, on l'a vu, qu'il est possible de répondre à cette question. On sait que la machine est susceptible d'adopter une stratégie féminine ou masculine au jeu de l'imitation. C'est par ce biais qu'une différence est établie entre

[279] Les références à l'homme pris au sens particulier sont fort nombreuses dans l'article: dans les trois premières pages de l'article seulement, il y en a neuf.

l'intérieur et l'extérieur du jeu, différence qui repose sur la notion de peau en tant qu'elle est porteuse d'une détermination sexuelle spécifique. Par ailleurs, la position de l'observateur extérieur doit, à la fin d'une partie, se confondre avec celle d'une machine. Bref, *le jeu de l'imitation est en lui-même une métaphore du corps rendue possible par l'identification de l'observateur extérieur à une machine possédant un statut sexuel absolument spécifique*.

On peut dès lors supposer que la position de l'observateur extérieur entretient avec le corps idéalisé de l'individu Turing un profond rapport de parenté. Plusieurs indices tendent à le montrer dans le texte. Ils ont tous rapport avec ce que Turing appelle, de façon ironique, le vice. A deux reprises, des expressions touchant à la perversité, au vice et à leur condamnation morale, reviennent sous la plume de Turing.

La première est celle qui a trait à l'exemple "pervers" choisi par Turing, dans lequel la machine ne produirait que des contradictions de type 0 = 1 [280]. On a vu que l'attitude qui consiste pour la machine à simuler l'ambiguïté dans ses réponses était considérée par Turing, sans doute de façon ironique, comme "perverse", dans la mesure où elle parodiait le registre du sexuel qui semblait lui être, par construction, étranger. Une autre expression employée par Turing va d'ailleurs dans le même sens. Il s'agit de l'expression : "mes vilaines habitudes" [«*my vicious ways*»]. L'expression de "vilaines habitudes" a, en anglais, une forte connotation sexuelle : elle désigne toutes les pratiques sexuelles de l'enfance ou de l'adolescence qui sont considérées comme "moralement répréhensibles". Pourquoi Turing utilise-t-il cette expression ? Il l'utilise comme une réponse à la question du manque d'originalité des machines. Turing écrit à ce propos [281]:

> **« Les machines me prennent très fréquemment par surprise. Cela est largement dû au fait que je n'effectue pas suffisamment de calcul pour décider ce à quoi je peux m'attendre de leur part ou plutôt parce que, bien que je fasse des calculs, je les fais de manière rapide et bâclée, en prenant des risques. Peut-être me dis-je : "Je suppose que le voltage doit être le même que là : de toute façon, on peut le supposer." Naturellement, je me trompe souvent et le résultat est pour moi une surprise car au moment où l'expérience a lieu, j'ai oublié ces hypothèses. Ces présupposés donnent matière à ce que je reçoive des leçons sur mes vilaines habitudes, mais ne jettent aucun doute sur mon honnêteté lorsque j'affirme que je fais l'expérience de ces**

---

[280] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 449.
[281] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 450-451.

surprises.»

Il n’y aurait donc pas de surprise si Turing ne se trompait pas dans ses calculs concernant les réglages physiques de la machine. On sait que cette erreur est le propre des êtres humains. Si aucune erreur n’était commise de sa part, son calcul *intellectuel* pourrait prévoir rigoureusement le réglage *physique* de la machine : dans ce cas, la synthèse entre le physique et l’intellectuel serait réalisée et le sexuel serait définitivement écarté. Le calcul, fonction majeure de toute machine discrète, relève donc bien d’un domaine qui n’est pas sexualisé. C’est seulement les erreurs imputables à Turing en tant que *personne* qui manifestent, comme on l’a vu en étudiant les stratégies de l’homme et de la machine, l’existence de la sexualité.

Remarquons à ce sujet que l’expression que nous commentons est “*mes* vilaines habitudes”, [*my vicious ways*]. C’est donc bien à lui-même que Turing attribue ces vilaines habitudes. Aussi peut-on voir ici une allusion à peine cachée à sa propre pratique sexuelle, l’homosexualité. L’identification entre le cas de la machine et celui de Turing s’étend donc jusque dans le registre sexuel : la machine est capable de suivre une stratégie masculine et féminine comme Turing est capable d’avoir une homosexualité. C’est donc bien de lui dont il est question ici, comme il en était déjà question auparavant quand il avait parlé de l’induction qui lui avait permis d’accéder au concept de machine universelle.

Il faut enfin remarquer que l’expression de “mes vilaines habitudes” apparaît dans un contexte particulier. Turing répond en effet à une objection qu’il met sur le compte d’une femme, Lady Lovelace, et qu’il appelle pour cette raison “l’objection de Lady Lovelace”. La comtesse Lovelace, qui était liée d’amitié avec Babbage, premier homme à avoir eu l’idée d’un calculateur universel, avait écrit un mémoire sur sa machine et avait fait remarquer dans la description qu’elle en donne, que cette machine ne pouvait rien inventer. C’est sur cette question de l’originalité du travail accompli par la machine que Turing répond en “s’adressant” à une femme, Lady Lovelace et parlant, à mots couverts, de sa pratique sexuelle “déviante”. Ce qui se trouve ici en jeu est donc le statut à accorder à la création. Il semble que, pour Turing, l’abandon de la norme sexuelle, c'est-à-dire de l'hétérosexualité, et la possibilité de l'engendrement biologique qui lui est lié, puisse être remplacé par une création "déviante" dont le résultat serait le concept de machine de Turing[282]. Le concept de machine de Turing semble donc entretenir avec la question de la détermination sexuelle de celui à qui on doit ce concept, Turing lui-même, des liens cachés.

Aussi est-ce maintenant de Turing et de son rapport à la sexualité qu'il faut parler.

3. Quelques souvenirs d'Alan Mathison Turing

Trois groupes de souvenirs doivent être évoqués pour rendre compte de ce qui fait le fond de l'article de 1950, l'identification personnelle de Turing à une machine. La question, on l'a vu, revient à celle de la construction de l'identité sexuelle de Turing. Le fait que Turing était homosexuel, qu'il fut condamné pour cette raison et que c'est sans doute l'un des motifs qui l'ont poussé au suicide en 1954 ne peut pas ne pas être pris en compte de ce point de vue. Les trois groupes de souvenirs peuvent être divisés chronologiquement : le premier a trait à la petite enfance de Turing, le deuxième a trait à son enfance, le troisième se rapporte moins à des souvenirs qu'à des comportements de sa vie adulte qu'il faut rapporter au passé pour qu'ils prennent tout leur sens.

31. La petite enfance de Turing

311. Le bannissement

Alan Turing est le fils cadet d'un fonctionnaire colonial, Julius Mathison

---

[282] Notons qu'en anglais, il y a plusieurs mots pour désigner la peau. Le plus fréquent est le mot de [*skin*] mais il en existe deux autres : [*pelt*] qui désigne la peau avec ou sans poil et [*hide*] qui est habituellement employé pour les animaux, sauf dans l'expression "sauver sa peau" qui est identique en anglais [*to save one's hide*]. On ne peut pas ne pas faire de rapprochement entre le substantif [*hide*] et le verbe [*to hide*] qui signifie "cacher" d'une part et ce que Turing cherche à "cacher" dans le jeu par rapport à la norme sexuelle d'autre part. Notons également que dans son étude de la morphogenèse, c'est précisément les taches sur la peau [*hide*] des animaux qu'il tentera d'expliquer en fournissant un modèle informatique de leur constitution. Cf. **A. M. Turing**, "The Chemical Basis of Morphogenesis", Phil. Trans. Roy. Soc. B 237, p. 24; republié dans [**A. M. Turing**, *Collected Works of A. M. Turing*, vol. 4, "Morphogenesis", North-Holland, 1992], p. 60. Il y a peut-être ici une réminiscence littéraire de la part de Turing : une des *Histoires comme ça* de Kipling décrit l'origine des rayures sur le pelage des animaux.

Turing, qui, après des études littéraires à Oxford, fit carrière en Inde, principalement à Madras. Julius Mathison était lui-même fils de John Robert Turing, le grand-père d'Alan, qui, après avoir fait des études de mathématiques à Trinity College à Cambridge, abandonna les mathématiques pour devenir pasteur. La mère de Turing, issue d'une famille de colons britanniques installée en Inde depuis le milieu du XIXème siècle, était née à Madras. Les deux familles étaient d'origine anglo-irlandaise et les parents de Turing se marièrent le 1er octobre 1907 à Dublin. Turing a été conçu en Inde mais il est né à Londres, le 23 juin 1912, son père ayant profité de ses vacances pour revenir en Angleterre avec sa femme et son fils aîné, John, qui était né le 1er septembre 1908 à Madras. Le père de Turing resta en Europe jusqu'en mars 1913, laissant ensuite en Angleterre sa femme et ses deux fils. Il ne revint en Angleterre que de façon épisodique, laissant jusqu'à trois ans d'intervalle entre ses visites. La mère de Turing rejoignit son mari en septembre 1913, laissant ses deux fils en nourrice à un couple de retraités, Monsieur et Madame Ward, qui habitaient St Leonards-on-Sea, un village en bordure de mer à côté de Hastings, sur la côte sud de l'Angleterre. Turing avait quinze mois. Il ne verra ses parents que par intermittence, durant toute son enfance. Il a lui-même dit que cette séparation précoce d'avec ses parents a eu une grande importance dans la construction de ses repères sexuels [283]. C'est aussi de cette époque qu'il date le souvenir de ces premières pratiques sexuelles avec les enfants du jardinier des Ward[284].

J'aurais tendance à interpréter cette première étape du développement psychologique de Turing comme une sorte de bannissement dans son propre pays, curieux effet en retour de la colonisation britannique de l'Inde. Le père de Turing apparaît comme un personnage lointain et de ce fait idéalisé. Hodges rapporte à ce sujet un souvenir d'enfance qui concerne indirectement Turing. Un jour, le frère aîné de Turing, John, demanda à son père ce qu'il détestait le plus. Celui-ci

---

[283] Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 132.
[284] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 77.

répondit sans hésitation : «La triche»[285]. On sait que le jeu de l'imitation fait du travestissement le but même du jeu : c'est sans doute une conséquence de l'idéalisation du père que d'avoir provoqué une inversion complète de point de vue sur la triche. Turing se voit condamné à faire de la triche une règle du jeu, lui pour qui la triche vise à cacher une forme de sexualité condamnable aux yeux du père.

Plus tard dans l'adolescence, Hodges rapporte, à propos de l'homosexualité dans les pensionnats anglais, les faits suivants [286]:

> **«Les contacts entre les garçons étaient chargés d'un potentiel sexuel : cela se reflétait dans les bannissements réels infligées aux liaisons entre garçons des différentes maisons[287] ou d'âges différents. Ces bannissements et les "racontars" ou le "scandale" avec lesquels ils étaient associés, ne faisaient pas *officiellement* partie de la vie de l'école privée, mais n'en étaient pas moins réels pour cela.. […]».**

Ainsi le bannissement de la petite enfance est-il peut-être associé à un bannissement dans l'adolescence, de nature sexuelle. Cette attitude de paria est sans doute une constante de la vie de Turing. On la retrouve plus tard, à l'extrême fin de sa vie, dans la dernière lettre qu'il envoya à R. Gandy en mars 1954[288]. Cette lettre était composée de quatre cartes postales dont la première portait le titre général de : "Messages du monde invisible" faisant sans doute référence au livre du physicien Eddington paru en 1929 qui s'intitulait *La science et le monde invisible*. Turing y faisait une analogie curieuse entre le principe d'exclusion de Pauli qui pose que deux électrons ne peuvent pas se situer au même endroit et le fait qu'à l'école, on interdisait aux garçons d'une même maison de se réunir comme ils l'entendaient, pour éviter des contacts qui auraient pu devenir "dangereux"[289]. A l'évidence, Eddington et Turing ne parlent pas du même du

---

[285] Le terme utilisé par le père de Turing, [*Humbug*], renvoie à la fois l'acte et à celui qui triche. Cf . **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 464 note.

[286] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 28.

[287] Dans l'enseignement secondaire anglais, les élèves étaient et sont encore regroupés en "maisons" qui comprend non pas une tranche d'âge correspondant à une classe mais une classe dans chaque tranche d'âge. Christopher et Alan n'appartenaient pas à la même maison.

[288] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., pp. 512-513.

[289] A. Hodges a reproduit ces cartes postales dans on livre. Le passage sur le principe d'exclusion est le suivant : «On ne pose le Principe d'Exclusion qu'au bénéfice des électrons eux-mêmes, qui

même “monde invisible”.

312. La circoncision

On doit dater de la même période, bien que ce ne soit pas précisé par Hodges, le fait que Turing ait été circoncis pour des raisons médicales[290]. Hodges rappelle à ce propos que Turing a dit qu’il avait par la suite conçu un grand regret de cette intervention et qu’il lui semblait qu’elle avait en partie déterminé son rapport à la sexualité[291]. La circoncision, d’acte médical devient acte symbolique de nature sacrificielle dans la mesure où elle apparaît bien comme une atteinte portée à l’intégrité de la peau, comme dans l’exemple du poème “Casabianca”, dans lequel le sacrifice de la peau - sa brûlure - était, pour le fils, l’offrande nécessaire pour envoyer, par-delà la mort, un signe de fidélité au père.

Hodges a retrouvé dans la correspondance de Turing une lettre de son enfance adressée à ses parents, datée du 11 février 1923 - Turing a onze ans - et qui mentionne le poème “Casabianca”[292]. L’allusion au poème est en rapport avec l’une de ses inventions : celle-ci consiste à construire un film sous forme d’un ruban de 16 images successives sur lesquelles étaient dessinées les images correspondant au récit. Mais le récit représenté n’est pas celui du poème de “Casabianca” mais celui d’une parodie de “Casabianca”. Il faut insister un instant sur cette parodie. Le poème de Casabianca a eu un curieux destin dans le système éducatif britannique : appris par cœur[293] par des générations d’écoliers jusqu’à aujourd’hui, on oublie généralement le poème original ainsi que le nom de l’auteur du poème, mais, en revanche, on fait, à partir du poème, de nombreux vers parodiques, souvent à consonance grossière ou grivoise[294]. La parodie du

---

pourraient se corrompre (et devenir [illisible] ou des démons) si on leur laissait la liberté de s’associer de façon trop libre».

290 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 77.

291 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 77.

292 «Chère Mère et Papa, (…) il y a 16 images pour chaque [film] et je me suis rendu compte que je pouvait dessiner ‘Le garçon se tint à la table à thé” vous savez la contine faite à partir de casabianca». **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 13.

293 Turing fait allusion à cet apprentissage par cœur dans “Computing Machinery and Intelligence” cf. **A. M. Turing**, “Computing Machinery and Intelligence”, op. cit., p. 457.

294 Je cite une parodie, parmi beaucoup d’autres :

poème telle qu'elle est rapportée par Turing "le garçon se tint à la table à thé" [*The boy stood at the tea table*] fait donc partie de la tradition écolière, que les parents de Turing connaissent certainement, ainsi que les connotations souvent sexuelles de ce genre de parodie, même si ce n'est pas celles-la que Turing rapporte dans sa lettre. Quoi qu'il en soit, la construction de ce film apparaît, dans le cas de Turing, comme mettant en rapport sa propre idée du sacrifice et la question de la sexualité, grâce à la construction d'une *machine* qui décompose en étapes successives représentées par des cases le récit parodié du poème. La peau fait donc elle-même signe par l'intermédiaire du ruban du film : retrouver l'intégrité de la surface de la peau passe par la création d'une machine qui expose le récit du sens donné à cette perte d'intégrité éprouvée dans la circoncision. Le fait que Turing se serve de l'exemple de "Casabianca" dans son article de 1950 montre la prégnance de ce poème et de ses parodies dans le sens qu'il attribue à la création de la machine de Turing[295].

On retrouve la mise en rapport de la peau et de la machine à l'époque suivante qui recouvre les souvenirs d'enfance de Turing.

32. L'enfance et l'adolescence de Turing

321. *Natural Wonders*

On lit dans la biographie de Turing qu'un livre marqua profondément l'enfance de celui-ci et contribua puissamment à éveiller sa vocation scientifique, comme il le confia lui-même plus tard à sa mère : il s'agit des *Merveilles naturelles que tout enfant devrait connaître* [*Natural Wonders Every Child Should Know*], ouvrage américain écrit par Edwin Tenney Brewster et publié à New York en 1912, que Turing reçut en cadeau à la fin de 1922 - il avait alors dix ans[296] -. Hodges reproduit plusieurs passages de ce livre pour enfants. Il commence par citer un extrait de l'avant-propos de l'auteur qui décrit les buts que celui-ci s'est

---

«Le garçon se tient sur le pont en feu
Ses pieds couverts de cloques
Il avait déchiré sa chemise de nuit jusqu'en bas
Et avait dû porter celle de sa sœur»

[295] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 457.

[296] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 11.

fixés [297]:

> **«Bref, il s'agit d'une tentative pour amener les enfants de huit ou dix ans, d'abord à poser et ensuite à répondre à la question : Qu'ai-je de commun avec les autres êtres vivants et comment est-ce que je m'en distingue ? De plus, j'ai essayé d'apporter au parent, embarrassé mais qui prend les choses aux sérieux, un point d'ancrage sur lequel il puisse lui-même se baser pour répondre à nombre des questions troublantes que tous les enfants posent - tout particulièrement la plus difficile de toutes : Par quel processus suis-je moi-même venu au monde ?».**

Après un chapitre intitulé "Comment le poussin se trouve dans l'œuf", le livre en arrive à un chapitre qui s'appelle "De quoi sont faits les petits garçons et les petites filles", chapitre qui reprend le titre d'une comptine anglaise. Brewster, en faisant référence à la comptine, fait remarquer [298] :

> **«Elle a cette part de vérité que les petits garçons et les petites filles sont loin d'être semblables et qu'il ne vaut pas la peine d'essayer de refaçonner l'un à partir de l'autre».**

Sans préciser la nature de cette différence, et après une diversion sur les œufs des étoiles de mer et des oursins, Brewster revient au corps humain [299]:

> **«Ainsi ne sommes-nous pas bâtis comme une maison de ciment ou de bois, mais comme une maison de brique. Nous sommes faits de petites briques vivantes. Quand nous grandissons, c'est parce que ces briques se divisent en demi-briques et finissent par reformer des briques complètes. Mais comment devinent-elles quand et où elles doivent grandir vite, quand et où elles doivent grandir lentement, quand et où elles ne doivent pas grandir du tout, voilà ce à quoi précisément personne n'a le moindre commencement de réponse».**

C'est sans doute parce que le corps humain est comparé par l'auteur de *Natural Wonders* à un mur de briques que Turing, dans "Computing Machinery and Intelligence", use de la même image quand il décrit la construction d'un individu à partir d'une cellule unique de peau. De même que dans *Natural Wonders*, la provenance de cette première cellule n'est pas précisée, de même c'est en suivant l'exemple de Natural *Wonders* que Turing identifie l'ensemble du corps humain à la construction d'un mur de briques.

Pour ce qui est de la provenance de la première cellule dans le processus de la génération humaine, Hodges fait remarquer que *Natural Wonders* ne donnait

---

[297] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 11. Non traduit dans la traduction française.

[298] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 12. Non traduit dans la traduction française.

[299] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 12

sur ce sujet aucun détail [300]. Comme dans *Natural Wonders* , cette absence d'origine se retrouve dans "Computing Machinery and Intelligence" : il n'est assigné aucune origine à la cellule de peau à partir de laquelle la reconstruction d'un individu entier est possible.

Pour ce qui est de l'identification de la construction du corps humain et de celle d'un mur de briques, il faut remarquer qu'elle repose sur une généralisation du cas de la cellule de peau à celui du corps humain dans son ensemble. L'usage que fait Brewster de l'image de la brique a sans doute conduit Turing à associer la constitution du corps humain en général à celle de la peau, dans la mesure où les cellules de peau se prêtent particulièrement bien à la comparaison avec un matériau de construction, puisqu'elles ont une structure stratifiée qui rappelle celle d'un mur de briques[301]. Le verbe qu'il emploie pour désigner cette construction à partir d'une cellule de peau, [*to rear*], dont la signification est "construire", "dresser" et aussi "éduquer", va dans ce sens. Cette construction artificielle à partir d'une cellule de peau a donc, de ce point de vue, une parenté très proche avec la "construction" naturelle du petit humain, dont il sera question à la fin de "Computing Machinery and Intelligence".

322. La première découverte touchant les réels calculables

Le premier succès mathématique de Turing a trait à la notion de réel calculable. Contrairement à ce à quoi on pouvait s'attendre, ce succès ne date pas de 1936 et de "On Computable Numbers …" mais de l'été 1927. Turing a quinze ans et sa médiocrité à l'école est telle qu'on parle de le rétrograder d'une classe. C'est à ce moment qu'il découvre seul la série infinie de la fonction tangente inverse :

---

[300] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 12.

[301] La comparaison des cellules de peau avec un mur de briques est encore employée aujourd'hui dans les pays anglo-saxons : «On a comparé la structure de l'épiderme à un mur de briques dans lequel les cellules riches en protéines (les briques) sont enfouies dans un environnement adhésif (semblable à un mortier) riche en lipides.» **D. Weedon ed.**,*Systematic Pathology*, Gen. Ed. W. St. C. Symmens, 3rd edition, Volume 9. *The Skin*, Churchill Livingstone, London, 1992, p. 3.

$$\tan^{-1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots$$

On sait combien l'étude des réels calculables et du rapport qu'ils instaurent entre discret et continu a eu une influence sur le destin intellectuel de Turing. On sait quelle parenté la différence entre le continu et le discret entretient avec la différence des sexes puisque Turing range la femme du côté du continu, l'homme du côté du discret et la machine, pour ainsi dire, “au milieu”. Rétrospectivement, cette première découverte de Turing apparaît comme capitale puisqu'elle manifeste l'orientation générale de la pensée de Turing.

Turing montre ce résultat à son professeur de mathématiques, absolument stupéfait, qui va en parler au professeur principal en qualifiant Turing de “génie”. Mais ce résultat n'éveille l'attention de personne d'autre et par la suite, même son professeur de mathématiques jugera son travail «pas très bon»[302]. On commence à le menacer d'expulsion. Au même moment, il contracte les oreillons et doit être placé en quarantaine. Sans suivre les cours, il décroche un prix de mathématiques et la menace d'expulsion est écartée.

Il semble que les mathématiques aient servi à Turing à se protéger d'une société dont il s'est senti très tôt rejeté, en partie sans doute pour des raisons d'ordre sexuel.

322. Christopher Morcom

A ce groupe de souvenirs qui se rapportent à l'éducation, on doit rattacher l'influence exercée par un camarade de classe de Turing, Christopher Morcom, que Turing a connu à partir du début de 1927.

Hodges rapporte qu'il s'agit du premier amour de Turing, amour non réciproque mais qui se développa en une réelle amitié. Bien qu'élèves de la même école, Christopher et Alan n'étaient pas censés se fréquenter, parce qu'ils n'appartenaient pas à la même “maison”. Christopher Morcom était d'un an plus âgé que Turing et s'intéressait, comme lui, aux questions d'ordre scientifique. C'est par le biais des mathématiques que Turing avait réussi à entrer en contact

---

[302] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 26.

avec Morcom[303] : Hodges rapporte que c'est à cette époque que Turing avait calculé les trente-six premières décimales de π et avait montré son travail à son ami, qui y avait trouvé une erreur et avait corrigé son résultat. C'est aussi avec lui que Turing commença à faire des expériences de chimie[304]. Notons enfin que Christopher Morcom était passionné par les jeux[305].

Turing a écrit un texte sur ses rapports avec Christopher Morcom. Deux traits caractéristiques y sont notés : le rapport de Morcom aux travaux intellectuels d'une part et à la morale de l'autre.

Turing note que le travail de Morcom était toujours meilleur que le sien [306]:

> **«Il était certainement très intelligent mais ne négligeait jamais les détails et par exemple faisait très rarement des erreurs de calcul […]. Chris affichait toujours une fierté charmante pour ses succès et je crois que c'était ce qui poussait l'instinct de compétition de chacun à faire quelque chose qui pourrait le fasciner et qu'il serait susceptible d'admirer. Cette fierté touchait également ses possessions. Il avait une façon d'afficher les vertus de son stylo de "recherche" qui me faisait saliver puis admettait qu'il essayait de me rendre jaloux».**

Si l'on se rappelle que c'est précisément l'exactitude "mortelle" qui caractérise la machine universelle et que c'est cette exactitude qu'il faut camoufler pour que la machine passe inaperçue au jeu de l'imitation, on voit combien l'identification de Turing à la machine universelle se fait par opposition mais aussi en souvenir de Christopher Morcom. Il y a bien opposition, parce que Turing a réussi, grâce au concept de machine universelle, à trouver le moyen de ne plus commettre de fautes d'inattention en arithmétique. Mais c'est aussi en

---

[303] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 35.

[304] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 36.

[305] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 43. Cette passion pour les jeux s'était transmise à Turing. Lors de son premier voyage aux Etats-Unis en 1937, Turing s'initia par l'intermédiaire de ses camarades de Princeton à des espèces de jeux de pistes appelés "Chasses au trésor" [*Treasure hunts*] qui consistaient, à partir d'indices composés d'anagrammes, de cryptogrammes ou de charades, à réunir un certain nombre d'objets. De retour en Angleterre, Turing pratiqua encore ce genre de jeux mais ceux-ci revêtirent une connotation sexuelle : les "messages" qu'il fallait décrypter étaient souvent à double entente, le deuxième sens n'étant compris que par ceux qui étaient "dans le secret", c'est-à-dire ceux pour qui l'homosexualité n'était pas un sujet tabou. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 397.

[306] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 36. Non traduit dans la traduction française.

souvenir de Morcom que Turing a pensé au moyen qui lui permettrait de dépasser le handicap des fautes d'inattention et d'égaler par ce biais les résultats de Christopher. C'est ce que montre le deuxième trait de la personnalité de Morcom sur lequel insiste Turing.

Turing décrit le sens moral de Christopher en ces termes [307]:

> **«Une chose au sujet de Chris que je crois être très inhabituelle était qu'il avait un code moral bien déterminé. Prenez les grossièretés par exemple. L'idée que Chris puisse avoir affaire à de telles choses semble risible et assurément, bien que je ne connaisse rien du tout de la conduite de Chris dans sa maison[308], j'aurais tendance à croire qu'il empêchait les grossièretés en faisant en sorte que les gens ne voulaient pas en prononcer plutôt qu'en faisant en sorte que les gens évite d'en dire pour ne pas le choquer».**

Si l'on se rappelle que les fautes de calcul sont décrites par Turing comme faisant partie de "ses vilaines habitudes", on voit comment celui-ci met en rapport, par le biais de Christopher, le domaine du calcul et celui de la morale. Si, d'un point de vue moral, les fautes de Turing sont liées à sa sexualité "déviante", en revanche ses fautes de calcul peuvent être, grâce au jeu de l'imitation, rapportées à l'humanité en général. Turing serait à la fois en mesure de faire *disparaître* ces fautes grâce à la machine universelle à laquelle il s'est identifié, tout en les *manifestant* au jeu de l'imitation, puisque la stratégie de la machine consiste précisément à faire des fautes de calcul. Le rapport à la morale est donc à la fois préservé et subverti, grâce à un étrange renversement : au lieu que ce soit, comme c'est le cas d'habitude, la morale qui soit affichée sans qu'elle soit suivie dans les faits, c'est l'inverse qui se produit puisque c'est la prétendue faute de la machine qui est affichée sans que celle-ci soit réelle. On voit comment le jeu de l'imitation permet à Turing de renverser sa propre attitude, lui qui se devait, selon les codes moraux en vigueur de son temps, de cacher ses "vilaines habitudes" en affichant un respect de la morale communément admise, c'est-à-dire en ne faisant que l'imiter[309].

---

307 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 37. Non traduit dans la traduction française.

308 Il s'agit de l'internat, au sens des maisons d'élèves.

309 Remarquons que les premiers essais d'automatisation de jeux par le biais de la machine universelle conduisirent Turing à concevoir une sorte de programmation aléatoire qui lui permettait d'écrire des "cadavres exquis" en forme de "lettres d'amour". En voici une, à peu près intraduisible : «*Darling Sweetheart, Your are my avid fellow feeling. My affection curiously clings*

Il est nécessaire de souligner un dernier point des relations de Turing avec Christopher Morcom. A la fin de leurs études secondaires, les deux amis passèrent un examen pour entrer à Trinity College, à Cambridge, qui était le collège du grand-père de Turing. Morcom fur reçu et Turing échoua. Mais Morcom n'eut pas le temps de tirer parti de cette admission à Trinity : il mourut le 13 février 1930, d'une attaque de tuberculose bovine qu'il avait contractée étant enfant en buvant du lait infecté. Ce décès eut une influence capitale sur Alan Turing. A plusieurs reprises après sa mort, Turing a écrit qu'il tenterait au cours de sa vie de suivre l'exemple de Morcom. Sa vocation scientifique est donc liée au souvenir de Christopher Morcom. Il écrit par exemple à la mère de Morcom [310]:

> **« […] Je suis assuré que je retrouverai Morcom quelque part et qu'il y aura du travail pour nous deux, de même que j'ai cru qu'il y en avait pour nous ici. Maintenant que je reste seul à le faire, je ne dois pas laisser tomber Chris mais au contraire je dois y mettre autant d'énergie, sinon d'intérêt, que s'il était encore ici. Si j'y parviens, je serai mieux armé pour apprécier sa présence que je ne le suis ici».**

Il s'agit donc pour Turing d'assumer *lui-même* la vocation scientifique de Christopher en tentant de réaliser ce qu'il n'a pas eu le temps de faire. Dès lors, la présence fantomatique de Christopher va occuper l'esprit de Turing. Il écrit à ce sujet un texte qu'il appelle "La nature de l'Esprit" qu'il envoie à la mère de Christopher [311]:

> **«Quant au lien entre l'esprit et le corps, je considère que le corps, du fait qu'il est un corps vivant, peut "attirer" et s'accrocher à un "esprit" et tant que le corps est vivant et éveillé, ils sont tous les deux étroitement unis. Quand le corps est endormi, je ne peux pas deviner ce qui arrive mais quand le corps meurt, le "mécanisme" du corps qui retient l'esprit s'en va et l'esprit trouve tôt ou tard un autre corps, peut-être immédiatement. Quant à la question de savoir pourquoi nous avons des corps, pourquoi nous ne vivons ou ne pouvons vivre libres comme des esprits et communiquer comme tels, nous pourrions sans doute le faire mais il ne nous resterait plus rien à faire. Le corps fournit à l'esprit de quoi s'occuper».**

Cette communication des esprits sans le support des corps semble avoir été inspirée à Turing par un autre de ses camarades de classe, Victor Beutell, qui,

---

to your passionate wish. My likings yearns to your heart. You are my wistful sympathy : my tender liking. Yours beautifully, MUC» Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 477.

[310] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 47.

[311] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., pp. 63-64.

outre sa foi chrétienne, croyait en la réincarnation [312]. Notons que cette communication des esprits ressemble fort au jeu de l'imitation. Il faut noter également une autre parenté avec l'article de 1950. Turing, dans la série d'objections qu'il s'adresse dans cet article, considère, peut-être avec ironie, l'objection de la télépathie qui viendrait fausser les règles du jeu, puisqu'une influence non-contrôlable de nature non-physique pourrait s'immiscer entre les joueurs et l'interrogateur. La réponse de Turing qui consiste seulement à dire qu'il faudrait placer les joueurs dans une pièce dont on pourrait s'assurer qu'elle n'est pas sensible aux perceptions extrasensorielles relève de la pure science-fiction. On pourrait dire qu'il y a là de l'ironie, mais cela n'explique évidemment rien. Le fait que Turing, après la mort de Christopher Morcom, ait cru en sa réincarnation[313], alors même qu'il faisait le projet de continuer l'œuvre scientifique de son camarade, semble plus proche de la vérité : c'est comme si la mort de Christopher lui donnait le moyen de s'identifier entièrement à lui, c'est-à-dire intellectuellement et physiquement. Pour ce faire, comme le veut le but du jeu de l'imitation, il faut avoir la possibilité de séparer l'esprit du corps pour qu'il soit capable de s'attacher à un autre corps qu'à son corps originel. L'objection de la télépathie renvoie aussi à l'exemple de Giacomo Casabianca qui, malgré la mort, s'était identifié à son père, et avait acquis, par ce fait, le rang d'exemple à suivre[314].

---

312 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 68.

313 On voit ici l'intérêt de ne pas traduire le mot anglais "implementation" par le néologisme "implémentation", comme nous le faisions remarquer plus haut. L'ambiguïté même du terme d'incarnation permet de conserver en français cette nuance de sens.

314 Je m'oppose donc ici à D. Hofstadter qui ne voit dans l'argument de la perception extrasensorielle qu'une espèce d'argument limite, dont il faut discuter le bien fondé scientifique. Celui-ci fait seulement remarquer : «Mon point de vue personnel, qui s'oppose à celui de Turing, est que la perception extrasensorielle n'existe pas». **D. Hofstadter**, *Gödel,Escher et Bach*, Interéditions, Paris, 1985, p. 671. Quelle que soit la nature *scientifique* du problème, la question ne me paraît pas se situer là. De même, J.-G. Ganascia, reprenant une par une les objections présentées par Turing dans "Computing Machinery and Intelligence", déclare que seule l'objection de la perception extra-sensorielle pose un véritable problème à l'intelligence artificielle : «Au terme de cet inventaire, il est amusant de constater que seule l'existence éventuelle de phénomènes parapsychiques a désarmé Turing. Tous les autres arguments ont pu être réfutés. Sauf à admettre l'intrusion de l'irrationnel dans le monde des phénomènes physiques, il n'existe donc pas d'objection de principe à l'idée d'une machine pensante. C'est là une conclusion majeure qui semble ne pas avoir été remise en cause depuis». **J.-G Ganascia**, *L'âme-machine; les enjeux de l'intelligence artificielle*, Le Seuil Paris, 1990, p. 210. L'aspect "amusant" de l'argument

Hodges note dans sa biographie que l'influence de Christopher Morcom sur Alan Turing ne dura pas après l'article de 1936 [315]:

> **«Il apparaîtrait bientôt comme un militant du point de vue matérialiste et se considérerait comme athée. Christopher Morcom mourait une deuxième fois et *Computable Numbers* marquait son décès. […] Christopher l'avait détourné du point de vue de *Natural Wonders* mais il y était revenu».**

Au contraire, il semble que la présence cachée de Christopher Morcom dans le jeu de l'imitation tel qu'il est exposé en 1950 va contre l'interprétation de Hodges. Contrairement à ce que celui-ci avance, il semble plutôt que jusqu'à la fin de sa vie, Turing ait prit au sérieux le problème du rapport entre la vision matérialiste de la science et l'apparente autonomie du spirituel, sans doute parce qu'il avait vu que c'était là que se jouait l'énigme de la sexualité dans son rapport à la constitution d'un domaine symbolique tel qu'il se manifeste dans la religion[316]. Un autre détail biographique va d'ailleurs dans ce sens. La mère de Christopher Morcom envoya à Turing, le jour où Christopher aurait eu vingt et un ans, c'est-à-dire, dans la tradition britannique, le jour de sa "majorité"[317], le stylo de Christopher, stylo de son invention qu'il appelait le "Stylo de Recherche" [*Research foutain pen*][318]. La mère de Christopher remit donc à Turing, le jour où son fils serait devenu un homme accompli, l'instrument qui permettait à ce dernier d'exposer ses recherches : il y a bien là un legs, à la fois matériel et

---

parapsychologique pose en fait le problème de la nature de l'esprit et ne me semble pas relever, comme J.-G Ganascia semble le supposer, de la problématique d'un "combat" du rationalisme scientifique contre les chimères psychologisantes de l'obscurantisme. On peut poser les termes du débat de façon plus "rationnelle" : quelle valeur cognitive attribuer à la croyance de Turing selon laquelle c'est maintenant lui qui incarne l'esprit de Christopher Morcom ? Si l'on accorde que, dans l'esprit de Turing, cette croyance a un sens, comment rendre compte du fonctionnement de cet esprit sans prendre en considération la croyance en question ? Que les individus considérés comme "normaux" puissent cependant avoir une vie spirituelle fantasmatique prenant part dans leur rapport au monde semble échapper aux deux auteurs que nous venons de citer.

315 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 108.

316 Dans l'une des dernières lettres de Turing, on trouve cet aphorisme : «La Science est une Equation Différentielle, la Religion est une Condition aux Limites». Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 513.

317 Il s'agit plus, en Angleterre, d'un rite de passage (accompagné d'une fête) que d'un age légal, comme le montre l'actuel abaissement du seuil de majorité à 18 ans, qui n'a pas changé le rite.

318 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 67.

spirituel, la majorité devenant un acte symbolique[319] par lequel chacun devient, grâce une transmission, responsable de sa vocation. C'est sans doute avec ce stylo que Turing écrivit "On Computable Numbers…" pendant l'été 1935. Notons enfin que c'est sur le bateau qui le ramenait à Princeton, en septembre 1937, que Turing perdit le stylo de Christopher [320]. Turing devint lui-même majeur le 23 juin 1933 et sa majorité est liée à deux événements. D'une part, c'est ce jour-là que Turing eut sa première relation homosexuelle suivie : il s'agissait d'une relation purement sexuelle et non sentimentale, au cours de laquelle tout ce qui avait trait à Christopher Morcom avait été laissé de côté[321]. D'autre part, c'est aussi à cette date que Turing retourna au domicile de Christopher pour une cérémonie du souvenir[322]. La majorité de Turing est donc liée à une double prise de responsabilité, liée à son "héritage" et à son penchant sexuel.

Abordons donc à présent quelques épisodes de la vie adulte de Turing.

33. Deux épisodes de la vie adulte de Turing

331. La science du décryptage

Turing, après la publication de "On Computable Numbers…" en 1937, fut consulté par les services secrets britanniques qui tentaient sans succès de déchiffrer les messages codés que l'Allemagne hitlérienne utilisait massivement. Cette collaboration devint, à partir de 1939, un travail à plein temps et Turing fut, comme un certain nombre d'autres universitaires, engagé par le service du chiffre[323]. Le déchiffrage des codes allemands allait occuper Turing pendant toute la guerre. Ce travail n'a pas dû manquer de lui rappeler que le décodage était un

---

319 Nous employons dans ce chpoitre le terme de "symbolique" dans son acception habituelle sans lui donner le sens qu'il a dans la citation que nous donnions de D. Widlöcher dans l'introduction de la deuxième partie.

320 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 167. Il n'est pas impossible qu'à cette époque, mais à cette époque seulement, Turing ait tenté de renoncer à l'"héritage" de Christopher Morcom, cette renonciation étant sans doute liée à la volonté d'atteindre une "majorité" dans sa vie sexuelle.

321 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., pp. 75-76.

322 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., pp. 75-76.

323 **F. H. Hinsley**, *British Intelligence in the Second World War*, vol I - VI, Her Majesty's Stationery Office Publications, London, 1979-1991.

de ses passe-temps favori lorsqu'il était encore à l'école, après qu'il eut gagné comme livre de prix un manuel de cryptologie[324].

331. 1. Le déchiffrage des codes allemands

Ce que l'on cherche à déchiffrer dans un message codé est moins le message lui-même que la *clé* qui permettrait de déchiffrer tous les messages. Comme on l'a déjà remarqué, le déchiffrage apparaît comme très semblable à la recherche d'un algorithme perdu qu'il faudrait reconstituer.

Le service britannique du chiffre était, à cette époque, une institution archaïque dont les méthodes n'avaient guère changé depuis la première guerre mondiale. En revanche, le bureau polonais du chiffre qui travaillait sur la même question, avait fait des progrès notables en renouvelant entièrement ses méthodes [325]. Le bureau polonais du chiffre avait engagé, dès le 1er septembre 1932, plusieurs mathématiciens parlant couramment l'allemand pour tenter de décrypter les messages chiffrés de l'armée allemande, codés grâce à une machine baptisée *Enigma*. C'est la conjugaison des compétences en mathématique - plus spécifiquement, en théorie des groupes - et en allemand qui permit à l'équipe polonaise de déchiffrer, avant tout le monde, les messages cryptés du Reich.

331. 2. L'*Enigma*

Cette machine, dont un modèle simplifié était vendu dans le commerce pour usage privé, comportait trois puis quatre rotors dont on pouvait intervertir l'ordre[326]. Sur chacun de ses rotors était fixé un anneau comportant les vingt-six

---

[324] **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 165.

[325] Cf. **M. Rejewski**, "How Polish Mathematicians Deciphered the Enigma", Annals of the History of Computing, Volume 3, Number 3, July 1981, p 213-234. L'article original a été écrit en polonais sous le titre "Jak matematycy polscy rozszyfrowali Enigme" dans les Annales de la Société Mathématique Polonaise, deuxième série, *Wiadomosci Matematyczne*, Volume 23, 1980, pp. 1-28.

[326] On trouve une description de l'Enigma dans **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 166-170, dans **M. Rejewski**, "How Polish Mathematicians Deciphered the Enigma", op. cit., p. 214-218 et dans **I. J. Good**, "Introductory Remarks for the article in Biometrika 66 (1979) "A. M. Turing's Statistical Work in World War II" publié dans [**A. M. Turing**, *Collected Works of A. M. Turing*, vol. 1 "Pure Mathematics", North-Holland, 1992], pp. 215-217.

lettres de l'alphabet. A chaque lettre était associé un contact électrique, contact qui était relié au rotor le plus proche. En pressant une touche du clavier de la machine à coder, semblable à un clavier de machine à écrire, le système de rotors encodait la lettre sur laquelle on avait appuyé (une impulsion électrique était envoyée à travers les rotors puis retournait à travers les rotors jusqu'à une planche où elle allumait une lampe électrique correspondant à une lettre) puis faisait tourner le premier rotor d'un vingt-sixième de tour. Chaque rotor tournait l'un après l'autre. Aussi, en réappuyant sur la même lettre du clavier de la machine à coder, la même lettre était-elle encodée différemment mais toujours selon un ordre, celui inscrit sur les rotors. Toutes les machines à coder de l'armée allemande possédaient la même configuration dans le câblage électrique des rotors, permettant ainsi le même encodage et le même décodage des messages[327].

L'originalité de la méthode utilisée par l'armée allemande à partir de Septembre 1938 était que, pour éviter que quiconque ne s'empare des instructions permettant de décrypter les messages, c'est-à-dire des instructions concernant l'ordre des rotors de la machine, cet ordre *n'était pas fixé à l'avance*. Chaque message devait donc commencer par les instructions - envoyées en double exemplaire pour pallier les mauvaises transmissions radio - permettant de placer les rotors dans l'ordre choisi pour la journée dans toutes les unités de l'armée allemande. Quelles étaient ces instructions ? Les premières instructions devait être données en clair puis étaient suivies par des instructions codées. C'est en exploitant ce nécessaire dédoublement que les mathématiciens polonais réussirent à retrouver la méthode permettant de constituer et de transmettre ces instructions.

Les instructions données en clair devaient être absolument aléatoires et laissées à la discrétion de l'encodeur. L'encodeur choisissait une certaine transformation pour une lettre donnée, par exemple a en x. Puis il choisissait une autre transformation, qu'il codait au moyen de la première. Il transmettait le tout au début du message, lui-même encodé à partir de la deuxième transformation. Ainsi la première transformation donnée en clair permettait-elle d'encoder la

---

327 Comme le fait remarquer Hodges, la conséquence d'un tel état de fait, était qu'une lettre ne pouvait pas être encodée par elle-même. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 168.

deuxième transformation qui, elle, servait à coder le message proprement dit.

331. 3. Les lettres féminines

Mais les mathématiciens polonais avaient découvert que le système permettant de passer de la première à la deuxième transformation laissait apparaître la *trace de l'ordre* des rotors. En effet, il arrivait quelque fois qu'une lettre encodée une première fois soit réencodée lors de la seconde transformation par la *même* lettre. Ainsi, si une lettre ne pouvait pas être encodée en elle-même *par construction*, en revanche une lettre encodée une première fois pouvait être encodée une deuxième fois de la même manière, c'est-à-dire sans changement apparent. Un tel cas de figure se produisait pour quarante pour cent des cas. C'est cette particularité que l'équipe polonaise avait repérée et exploitée.

Leur méthode reposait donc sur la découverte de répétitions des lettres identiques dans l'encodage des instructions codées au début de chaque message. A partir de ces répétitions fortuites, il était possible, si l'on possédait de grandes quantités de messages codés, de reconstituer l'ordre relatif des rotors adopté ce jour-là par toutes les *Enigmas* de l'armée allemande. Le caractère apparemment aléatoire du message laissait donc le moyen de reconstituer la nécessité de la clé, pourvu que l'on possède une quantité suffisamment large de messages où retrouver l'indice de la répétition des lettres. On voit combien ce schéma ressemble à celui du jeu de l'imitation dans lequel Turing assure que la reconstitution de la clé, l'identité sexuelle, ne peut pas se produire. Un détail montre d'ailleurs la proximité du schéma du déchiffrage et de celui du jeu de l'imitation, dans l'esprit de Turing.

A. Hodges fait remarquer que, dans la méthode de déchiffrage, la lettre répétée était, «sans raison apparente», appelée "féminine"[328]. Cet usage ne peut pas manquer d'éveiller l'attention, quand on sait que Turing allait écrire en 1950 "Computing Machinery and Intelligence" : le jeu de l'imitation ne pouvant se jouer que par rapport à la différence des sexes, cet usage sexualisé des lettres demande à être explicité.

---

[328] **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 173.

A ma connaissance, ni Rejnewski, ni Hinsley[329] n'attestent cet usage. En revanche, on trouve un certain nombre d'indices permettant d'en reconstituer les raisons dans les souvenirs que I. J. Good, qui faisait partie de l'équipe de Turing, retrace à l'occasion de la traduction en anglais de l'article écrit en polonais par Rejnewski qui, pour sa part, faisait partie de l'équipe polonaise.

331. 4. Une nouvelle unité statistique : le Ban

Après l'invasion de la Pologne par l'Allemagne, l'équipe polonaise attachée au service du chiffre était passée en France. Elle avait continué, après la défaite française, à travailler en zone libre au profit des Anglais, grâce au chef du service français du chiffre. Les Anglais avaient installé leur propre service du chiffre hors de Londres, à Bletchley Park, pour éviter les bombardements.

Ils avaient repris à leur compte, mais à une autre échelle, les méthodes des Polonais et tout particulièrement le classement des lettres dites féminines qui permettait de reconstituer l'ordre relatif des rotors de la machine à crypter. Le classement était opéré par un système de cartes perforées. I. J. Good rappelle que Turing à Bletchley Park fut le premier à inventer une méthode statistique permettant d'accorder aux “configurations féminines” apparaissant dans le stock de cartes perforées une plus ou moins grande importance selon leur fréquence. L'unité de mesure de cette fréquence avait été baptisée par Turing le “ban”. Voici ce qu'en dit I. J. Good [330]:

> **«Les cartes perforées des messages étaient appelées feuilles de Banbury, parce qu'elles étaient imprimées dans la ville de Banbury. Les perforations étaient laborieusement faites par des équipes de jeunes femmes que l'on appelaient “les filles”. Les configurations qui se répétaient étaient comptées en utilisant des “poids d'évidence” (des logarithmes de facteurs de Bayes) mesurés en “décibans”. Les noms de *déciban* et de *ban* avaient été inventé par Turing.»**

On peut grâce aux souvenirs de Good, deviner pourquoi cette unité de mesure avait été appelée par Turing le “ban”. Les cartes perforées utilisées pour le classement des lettres étant imprimées dans la ville anglaise de Banbury, on peut en déduire, bien que ce ne soit pas expressément mentionné par Good, que c'est

---

[329] Au moins dans les deux premiers volumes, publiés en 1979 et 1981, les tomes suivants (tomes 3 à 6) n'ayant pas encore été publiés quand Hodges a écrit sa biographie de Turing.

[330] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 233.

l'une des raisons pour laquelle Turing appela l'unité de mesure statistique inventée par lui le "ban". Aussi l'équipe de Bletchley Park appelait-elle la procédure permettant la comparaison statistique des fréquences dans les configurations féminines un "Banburisme"[331]. Notons que le terme de "Banbury" aura aussi, après la guerre, une autre fonction : Turing décomposera le nom et appellera "Bury" et "Unbury" (en laissant tomber le "Ban") certaines procédures de programmation servant de sous-programmes[332].

331. 5. Le "Banburisme"

Cette procédure exploite la cohérence logique du système d'encodage, c'est-à-dire tente de mettre au jour les possibilités de contradiction qui peuvent apparaître dans les combinaisons de lettres pour une même position des rotors[333].

Une fois qu'un certain nombre de configurations féminines ont été découvertes et qu'un mot probable en allemand a été déterminé dans un stock relativement peu élevé du fait de l'aspect stéréotypé des messages militaires, on peut éliminer un grand nombre de possibilités de permutation en découvrant des contradictions entre des assemblages de lettres : une même lettre cryptée ne peut pas, pour une position donnée des rotors, être transformée à la fois en deux lettres différentes. Donc toutes les combinaisons construites à partir de cette position des rotors sont éliminées. Il faut alors construire une machine qui simule ces implications et qui *s'arrête* sur les contradictions. Cette machine, en souvenir du travail des membres l'équipe polonaise, portait le même nom que la leur : la "Bombe". Turing et un membre de l'équipe, Gordon Welchman, parvinrent alors

---

[331] **I. J. Good**, à la suite de l'article de **M. Rejewski**, "How Polish Mathematicians Deciphered the Enigma", p. 233 : « C'était en rapport avec le Banburisme que Turing eut un certain nombre d'idées statistiques nouvelles ou relativement nouvelles, telles que l'analyse séquentielle ou la forme non triviale de l'opération empirique de Bayes.» Un autre système de codage que celui fondé sur l'Enigma et qui avait été baptisé "Fish" fut décodé grâce aux idées statistiques propres à Turing. La procédure s'appelait en anglais le "Turingismus". Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., pp. 230-231.

[332] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 343 et **A. M. Turing**, "Proposal for Development in the Mathematics Division of an Automatic Computing Engine (ACE)" , reprint dans [**B. E. Carpenter** et **R.W. Doran eds.,** *A. M. Turing 's ACE Report of 1946 and other Papers*, MIT Press, Cambridge, 1986], p. 76.

[333] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 183.

à réaliser un “tableau diagonal”, composé de deux fois vingt-six lettres et dans lequel chaque liaison entre lettres (suscitée par les répétitions féminines indiquant des combinaisons entrant dans un mot probable) est établie par une connexion électrique. Ce tableau, relié à la “Bombe”, permet d’étudier les contradictions entre combinaisons et d’éliminer celles-ci. Il s’agit donc d’étudier les conditions logiques d’arrêt d’une machine. Dans le cas où le choix du rotor simulé est correct, deux solutions sont envisageables quand la “Bombe” s’arrête : soit le tableau de connections est correct et il n’y a aucune contradiction, c’est-à-dire que les connections matérialisées dans le tableau sont celles du rotor; soit il y a une contradiction et les liaisons entre les lettres du tableau sont toutes mauvaises et indiquent de ce fait les bonnes : il n’y a pas besoin d’étudier une par une chaque liaison et le travail avance vingt-six fois plus vite.

Le travail de l’équipe dont Turing faisait partie a donc consisté à relier, par un procédé mécanique, les places physiques des rotors et les traces de la langue naturelle, en établissant par une analyse statistique leur correspondance.

On voit ici la parenté avec un aspect du jeu de l’imitation : de même que le déchiffrage des messages cryptés consiste à mettre en correspondance des données linguistiques (les répétitions écrites puis comptabilisées) et des caractéristiques physiques (le câblage spécifique à un ordre des rotors) de même dans le jeu de l’imitation, l’interrogateur essaye d’établir une correspondance entre les réponses cryptées qui lui sont fournies et la nature physique de la personne interrogée. De ce point de vue, le fait d’avoir appelé les lettres répétées des lettres “féminines” a pu aussi servir à Turing lors de l’invention du jeu puisque la stratégie de la femme consiste précisément à répéter son identité, cette répétition finissant par renvoyer à sa nature physique. La question de la possibilité d’un mensonge de la part de la femme avait été évoquée par Turing à cette époque : à la fin de la guerre, Turing fut chargé construire une machine servant à crypter la voix humaine. Ayant offert un prix à qui trouverait un nom pour la machine, c’est celui de “Dalila”, imaginé par Robin Gandy, qui emporta son adhésion parce qu’elle “avait trahi les hommes”. On sait que cette “trahison” sera finalement le propre de la stratégie de la machine.

La détermination physique et sexuelle était donc *déjà présente* dans le travail de décryptage[334].

332. Joan Clarke

Joan Clarke était une mathématicienne qui avait été engagée pendant la guerre par le bureau du chiffre et qui habitait à Bletchley Park avec tous les autres chercheurs. Turing et elle se sont connus à partir du printemps 1941. Turing la présenta à sa famille. Hodges rapporte que Turing était content du fait qu'il pouvait lui parler "comme à un homme"[335]. Il rapporte d'autre part que Turing avait émis le souhait d'avoir des enfants. C'est à cette époque que Turing développa son goût pour les questions de morphogenèse, questions qui allaient l'occuper dans ses derniers articles[336]. Turing a rompu avec Joan Clarke lors d'un congé pendant lesquels ils étaient allés sur le lieu des vacances que Turing avait passées en famille en 1916 [337]. Hodges rapporte[338] qu'il cita, lors de la rupture, les dernières lignes de la *Ballade de la geôle de Reading* d'Oscar Wilde qui, lui aussi, avait, pour des raisons touchant à son homosexualité, été en butte aux blâmes de la morale victorienne. Aux rares personnes qui connaissaient leurs fiançailles, Turing prétexta d'avoir eu en rêve le sentiment que sa fiancée ne serait pas acceptée par sa famille. Leurs relations se poursuivirent sur un mode amical après leur rupture.

Un passage de l'article de 1950, à résonance ironique, peut être relié à cette série de faits portant sur le mariage. Dans l'objection que Turing appelle l'objection théologique"[339], celui-ci veut montrer que c'est *autant* un péché de créer des machines intelligentes que de procréer de façon naturelle. Turing associe donc la procréation naturelle et la construction artificielle par le biais d'un interdit

---

334 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence,* op. cit., p. 273.

335 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 206.

336 **A. M Turing**, (1952). "The Chemical Basis of Morphogenesis", Phil. Trans. Roy. Soc. B 237 (1952), republié dans [**A. M. Turing**, *Collected Works*, Morphongenesis, ed. P. T. Saunders, vol. 4, North-Holland, Amsterdam, 1992], p. 1- 40. Turing considérait cet article comme aussi important que "On Computable Numbers…".

337 Précisément à Kimelfort dans les Western Highlands.

338 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 217.

339 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 443.

de nature théologique. Pour asseoir sa thèse, il lui faut donc tout d'abord faire remarquer que c'est *déjà* accomplir un péché que de procréer de façon naturelle [340]:

> **«En essayant de construire de telles machines nous ne devrions pas plus usurper irrévérencieusement Son pouvoir de créer des âmes que dans les circonstances de la procréation des enfants : nous sommes plutôt, dans les deux cas, des instruments de sa Volonté fournissant une demeure pour les âmes qu'Il crée».**

Même la procréation naturelle (et pas seulement la construction artificielle) revient à manquer de respect envers Dieu : dans ces conditions, toute création dont l'attribution ne serait pas rapportée à sa source divine serait de l'ordre du péché. Plutôt que d'avoir à assumer ce péché, il vaut mieux, dit Turing, considérer que le cas de la procréation naturelle et le cas de la construction artificielle sont identiques : il ne s'agit pas alors d'une création véritable mais seulement de la transmission temporaire d'une enveloppe matérielle à une entité déjà créée et dont la création n'a pas à être assumée.

L'ironie consiste donc pour Turing à se placer du point de vue d'une autre objection qu'il a critiqué auparavant et qu'il a appelée "l'objection de l'autruche"[341]: l'objection de l'autruche consistait à dire que les conséquences d'une pensée mécanique seraient tellement affreuses qu'il vaut mieux renoncer à une telle hypothèse. L'argument "théologique" de Turing est identique : plutôt que d'assumer le risque de remettre en question l'interdit théologique sur la procréation, il vaut mieux croire que cette procréation, naturelle ou artificielle, n'existe pas[342].

Mais cette ironie laisse intacte l'idée qui a permis d'envisager d'un même point de vue le cas de la création naturelle et le cas de la création artificielle, idée qui consiste à considérer que la création est de l'ordre du péché et qu'il y a là un interdit de nature théologique. Aussi la création et l'invention sont-elles toutes les deux considérées par Turing comme de simples transformations : l'objection dite de "Lady Lovelace" qui porte sur le fait que les machines n'inventent pas, va

---

[340] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 443.
[341] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 444.
[342] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 443.

donc de pair avec l'objection dite "théologique" qui fait remarquer que la création n'existe pas. La première objection porte sur l'invention intellectuelle et la seconde sur la création physique : toutes les deux sont référées à un interdit sur la notion d'engendrement, matériel ou intellectuel.

La volonté de création doit donc utiliser des biais détournés pour s'exprimer. Or s'il existe un interdit sur la création intellectuelle - qui a lieu dans l'intériorité de l'esprit - et matérielle - qui a lieu dans l'intérieur du corps -, il n'y a plus que sur *lui-même* en tant qu'il possède une peau qui n'est ni à l'intérieur ni à l'extérieur, que peut se porter cette volonté de création, dans une sorte d'auto-engendrement dont on peut retrouver les traces dans "Computing Machinery and Intelligence". C'est ce qu'il nous faut étudier maintenant.

4. L'auto-création

L'article de 1950 a ceci de particulier qu'il décrit non seulement des épisodes autobiographiques mais qu'il remonte jusqu'à la naissance imaginaire de Turing lui-même, naissance confondue avec son identification à la machine universelle : aussi la dernière partie de l'article qui s'intitule "Machines qui apprennent", est-elle une description du mode d'apprentissage d'une machine-enfant[343] qui possède un certain nombre de traits de ressemblance avec l'identification de Turing à la machine universelle.

41. La peau de l'esprit

La description de la naissance est, au départ, référée à la fois au domaine physique et au domaine intellectuel. Turing se sert pour ce faire d'une description qui renvoie explicitement à l'image de la peau [344]:

> **«L'analogie de la peau de l'oignon est aussi utile. En considérant les fonctions de l'esprit ou du cerveau, nous trouvons certaines opérations que nous pouvons expliquer en termes purement mécaniques. Nous disons que cela ne correspond pas à l'esprit véritable : c'est une sorte de peau que nous devons arracher si nous voulons trouver l'esprit véritable. Mais dans ce qui reste, nous trouvons une autre peau à arracher et**

---

[343] On a déjà noté que les ingénieurs de Manchester appelait leur premier ordinateur la "machine-bébé" à cause de la petite taille de sa mémoire. Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 385.

[344] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., pp. 455-456.

ainsi de suite. En procédant de la sorte, arrivons-nous jamais à l'esprit "véritable" ou parvenons-nous finalement à la peau qui ne contient rien ? Dans ce dernier cas, tout l'esprit est mécanique (ce ne serait pas une machine à états discrets, cependant. Nous en avons discuté.)».

On remarque que Turing commence par associer l'esprit et le cerveau mais qu'il restreint ensuite son attention au seul cas de l'esprit, c'est-à-dire au seul domaine intellectuel. Cette restriction est semblable à celle que Turing opère dans le choix des machines, quand il justifie le fait de ne prendre en considération que les ordinateurs digitaux.

L'esprit est donc assimilé à une peau d'un genre bien particulier puisque celle-ci ne distingue pas l'intérieur de l'extérieur : il s'agit d'une peau qui n'a plus de contenu. Pour réussir à faire en sorte que l'*intériorité* soit exposée sur le même registre que celui de l'*extériorité*, il faut concevoir la peau comme une simple surface, en supprimant son rôle premier qui est de séparer l'intérieur du corps de l'extérieur. Cette suppression exige un *dépiautage* jusqu'à une peau primitive qui serait l'esprit lui-même. L'esprit peut alors, dit Turing, être identifié à une machine. Mais les caractéristiques de cette machine sont curieuses car qu'est-ce donc qu'un esprit qui serait assimilé à quelque chose qui serait *à la fois* mécanique et non-discret ? *Il s'agit d'une machine*-o. Turing ne répond pas à la question de savoir comment il serait possible de caractériser cette machine mais fait seulement semblant d'y répondre en prétendant y avoir *déjà* répondu. Or cette réponse n'existe nulle part dans l'article, au moins explicitement. En revanche, l'identification de la peau à une machine a déjà eu lieu dans "Computing Machinery and Intelligence" et ce, tout au début de l'article, dans ce que l'on a appelé "l'objection de l'équipe d'ingénieurs". Il est remarquable de constater que la question de l'apprentissage des machines entretient avec cette objection d'étroits rapports de parenté.

42. La peau des machines

La description des modes d'apprentissage des machines commence en effet par quelques remarques sur l'éducation en général. Turing fait remarquer que l'état dans lequel se trouve un esprit adulte repose sur trois types de composantes [345]:

> **«(a) L'état initial de l'esprit, par exemple à la naissance,**
> **(b) L'éducation à laquelle il a été soumis**
> **(c) D'autres expériences auxquelles il a été soumis et que l'on ne doit pas décrire comme relevant de l'éducation».**

Ces trois conditions entretiennent avec les trois conditions utilisées par Turing pour décrire ce que l'on a appelé "l'objection de l'équipe d'ingénieurs" un rapport étroit.

Premièrement, en effet, dans cette objection, la naissance par voie naturelle n'était pas considérée comme pouvant entrer dans le cadre d'une naissance mécanique : ici, au contraire, la naissance biologique est considérée comme ayant un "état initial", c'est-à-dire qu'elle obéit au schéma de fonctionnement de la machine universelle qui possède, elle aussi, un état initial.

Deuxièmement, Turing observe que dans le processus d'éducation, il était important de remarquer que l'éducateur n'était pas tenu de savoir précisément ce qui se passait *à l'intérieur* de la machine [346]: Turing reproduit donc ici mot pour mot ce qu'il avait déjà fait remarquer à propos de l'équipe d'ingénieurs.

Troisièmement enfin, l'esprit de l'adulte est constitué en partie par des "expériences" que l'on ne *doit* pas décrire comme relevant de l'éducation. La forme impérative de la troisième condition semble donc associer ces "expériences" à des faits que l'on ne doit absolument pas rapporter à l'éducation et l'on peut supposer qu'il s'agit ici d'une référence à peine cachée à la sexualité. De même, dans l'objection de l'équipe d'ingénieurs, le statut de la peau était laissé dans l'ombre, sans que l'on sache vraiment s'il fallait ou non lui accorder un statut d'ordre sexuel.

La description de cette machine-enfant entretient donc des rapports profonds avec l'équipe d'ingénieurs, qui pouvait, comme le faisait remarquer Turing, se limiter à un seul membre masculin. Il faut étudier ces rapports. La question de la construction d'une machine qui apprend exige donc d'être divisée en trois étapes. Quelles sont ces trois étapes ? La première doit réussir à résoudre

---

345 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 455.
346 **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., pp. 458-459.

la question du type de programme capable de simuler le cerveau de l'enfant à la *naissance*. La seconde doit tenter de montrer comment s'effectue *l'apprentissage* lui-même. La troisième doit rendre compte des *expériences* que l'on ne doit pas rapporter à l'éducation au sens strict du terme. Étudions ces trois étapes.

42. 1. La naissance de la machine-enfant

Dans "Intelligent Machinery", Turing avait fait remarquer que la "machine de papier", c'est-à-dire l'homme muni seulement d'une liste d'instructions écrites, d'un papier, d'un crayon et d'une gomme, était en fait une machine universelle. Le problème qui restait à résoudre consistait à montrer qu'il était possible de *devenir* une machine universelle. En filant la même métaphore, la comparaison que Turing utilise dans "Computing Machinery and Intelligence" pour exposer son point de vue sur la naissance de la machine-enfant renvoie à la page d'écriture et, par contrecoup, à la peau [347]:

> **«On peut supposer que le cerveau de la machine-enfant est comme un carnet que l'on achète en papeterie. Plutôt peu de mécanisme et beaucoup de feuilles blanches (Mécanisme et écriture sont, de notre point de vue, presque synonymes). Notre espoir est qu'il y ait si peu de mécanisme dans le cerveau de l'enfant, qu'il soit facilement programmable».**

L'esprit ayant été au préalable assimilé à une peau vide de tout contenu, dans l'analogie avec la peau d'oignon, la surface de la peau et la surface de la page se trouvent donc associées et c'est ce qui permet d'identifier l'esprit de l'enfant à une machine. Le programme susceptible de simuler le cerveau de la machine-enfant est donc écrit sur une surface très particulière. Il faut se rappeler de surcroît qu'il s'agit ici de la description d'une naissance : de quel enfant s'agit-il puisque cette naissance a partie liée à l'identification à une machine universelle ? Il s'agit de celle de Turing *lui-même*. L'invention de la machine universelle est donc assimilée par Turing à une naissance rendue possible par le biais de l'écriture sur la peau : c'est ce que Giacomo Casabianca accomplit en sacrifiant son corps et la différence que celui-ci institue entre le contenant et le contenu. La

[347] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 456.

façon dont Turing décrit sa propre naissance en tant que machine universelle est donc semblable à la confection d'un tatouage. On peut donc dire que Turing s'auto-engendre par l'invention de la machine universelle et que l'expression de cet auto-engendrement est le tatouage de sa peau par l'écriture[348]. Le *dépiautage* nécessaire à la découverte de la peau primitive qu'est l'esprit est donc suivi d'un *tatouage* de cette même peau.

42. 2. L'éducation de la machine-enfant

En quoi consiste l'éducation proprement dite de la machine-enfant ?

Dans "Intelligent Machinery", Turing avait expressément fait remarquer que l'éducation d'un enfant consistait à devenir une machine universelle, c'est-à-dire de passer d'un état inorganisé à un état organisé [349]:

> **«Tout ceci suggère que le cortex d'un nourrisson est une machine inorganisée, qui peut être organisée au moyen d'un entraînement adéquat. Cette organisation pourrait provoquer une modification de la machine qui résulterait dans la constitution d'une machine universelle ou de quelque chose approchant. […]. Il est intéressant de faire des expériences avec des machines inorganisées admettant des types définis d'interférence et d'essayer de les organiser, par exemple de les modifier pour en faire des machines universelles».**

Comment s'opère cette éducation ? Elle possède deux volets : l'assujettissement à une discipline et la possibilité d'une initiative. Pour Turing,

> **«Transformer un cerveau ou une machine en une machine universelle est la forme la plus extrême de la discipline. Sans quelque chose de cette nature, il ne peut pas y avoir de communication véritable. Mais la discipline n'est certainement pas suffisante pour produire de l'intelligence. Ce qui est requis en plus est appelé initiative. Cet énoncé devra servir de définition. Notre tâche est de découvrir la nature de ce résidu tel qu'il apparaît en l'homme, et d'essayer de le copier dans les machines»[350].**

On sait que dans "Computing Machinery and Intelligence", Turing avait donné une exemple de cette "forme extrême de discipline" avec l'exemple du sacrifice corporel de Giacomo Casabianca. Quelle est la nature du "résidu" qu'est

---

[348] La dernière partie de "Computing Machinery and Intelligence" s'achève donc par la "naissance" de Turing et l'article doit donc se lire comme un *acte de réminiscence* : plus on s'avance dans la lecture de l'article et plus on recule dans le temps.

[349] **A. M. Turing**, "Intelligent Machinery", op. cit., pp. 43-44.

[350] **A. M. Turing**, "Intelligent Machinery", op. cit., p. 49.

l'initiative ? Elle correspond à la dernière peau qui ne recouvre rien dans l'analogie de la peau d'oignon du jeu de l'imitation. Dans "Intelligent Machinery", elle apparaît de façon typique sous la forme des problèmes de nature computationnelle [351]:

> **«Une sorte de problèmes tout à fait typique requérant quelque initiative consiste en ceux de la forme "Trouver un nombre *n* qui ". Cette forme recouvre une grande variété de problèmes de la forme "Tâchez de trouver un moyen de calculer la fonction … qui nous permettra d'obtenir la valeur pour les arguments … avec telle précision … pendant un temps … en utilisant une machine universelle …"; ces problèmes sont réductibles à cette forme, car le problème est évidemment de trouver un programme pour la machine en question et il est facile de mettre les programmes en correspondance avec la suite des entiers positifs de telle sorte que, soit que l'on donne le nombre soit que l'on donne le programme, l'autre terme puisse être trouvé aisément. Nous ne nous tromperons guère pour le moment si nous faisons la supposition que tous les problèmes sont réductibles à cette forme. Il sera temps de penser autrement quand quelque chose apparaîtra qui, à l'évidence, ne sera pas de cette forme».**

Turing considère ainsi comme relevant de "l'initiative" toute solution computationnelle à un problème arithmétique : il écarte ce faisant non pas les problèmes qui ne seraient pas susceptibles de recevoir un traitement computationnel (puisque ces problèmes ne sont pas définissables *a priori*) mais *l'idée* d'un problème qui résisterait à tout traitement computationnel. Or la constitution de la notion formelle de machine universelle était liée à une telle idée, puisque la solution que Turing avait lui-même apporté au problème de l'arrêt en manifestait la légitimité. Cette idée est donc bien *écartée* sans qu'on sache exactement pourquoi[352]: il y a ici comme un "résidu du résidu" qui ne serait pas accessible à un traitement computationnel sans que ce résidu soit clairement reconnu comme tel.

Que ce soit dans le cas des machines pour lequel on se heurte à l'existence inassignable d'un "résidu de résidu" ou que ce soit dans le cas des être humains où l'on se heurte à l'existence d'une peau terminale ne recouvrant rien et qui

---

[351] **A. M. Turing**, "Intelligent Machinery", op. cit., p. 200.

[352] C'est ce que notait Hodges quand il faisait remarquer : «Curieusement, il [Turing] n'était pas intéressé par sa propre découverte touchant l'existence de questions absolument indécidables à l'intérieur du cadre du modèle de la machine à états discrets, mais il se souciait beaucoup plus de mettre l'accent sur le pouvoir de la machine universelle.» Cf. **A. Hodges**, "Alan Turing and the Turing Machine" dans [**R. Herken ed.**,*The Universal Turing Machine,* a half-century survey, Oxford Science Publication, Oxford, 1988], pp. 10-11.

serait celle d'un esprit à la fois mécanique et non-discret, on rencontre la même aporie touchant la constitution de la notion d'intelligence. C'est cette aporie que nous avions appelée le "problème du continu", car il apparaît comme le cœur de la théorie computationnelle de l'esprit.

On sait que c'est par un appel à la notion d'aléatoire que Turing tente de prendre en compte ce problème. Il use, pour exposer son point de vue, d'une comparaison entre la possibilité d'un apprentissage et la théorie évolutionniste des espèces. La prise en compte d'un élément aléatoire est indispensable dans le cas de l'évolution, parce qu'il n'y a pas moyen de stocker en mémoire les différents choix génétiques déjà utilisés par les espèces[353]. Turing considère alors que, de même, il serait utile d'inclure un élément aléatoire pour rendre possible l'apprentissage de la machine-enfant. La machine-enfant peut en effet avoir à effectuer des transformations dans son propre programme, selon le type de problèmes qu'elle doit résoudre. Cette transformation renvoie-t-elle exclusivement à une transformation dans le *programme* de la machine-enfant ou aussi à une transformation dans sa constitution *matérielle* ? "Intelligent Machinery" reconnaissait l'existence du problème et déclarait ne traiter que de la transformation du programme[354].

En fait, on sait que cette ambiguïté renvoie au critère même du jeu de l'imitation, celui de la différence entre le physique et l'intellectuel tel qu'elle émane de la différence des sexes : c'est pourquoi il faut maintenant essayer de rendre compte de la troisième étape du processus éducatif, à savoir ces "expériences" que l'éducation ne *doit* pas prendre en considération.

42. 3. Les expériences de la machine-enfant

Certaines expériences doivent être rapportées au processus éducatif tandis que d'autres n'en relèvent pas directement. Étudions ces deux sortes d'expériences.

Pour Turing, l'éducation d'un enfant consiste à intervenir de l'extérieur

---

[353] Cf. **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., pp. 459- 460.

[354] Dans "Intelligent Machinery", Turing distingue la "transformation par le papier" de la "transformation par le tournevis". Cf. **A. M. Turing**, "Intelligent Machinery", op. cit., p. 37.

sur son fonctionnement interne par le biais de stimuli physiques. Dans "Intelligence Machinery", Turing fait remarquer qu'il est difficile de construire une machine qui reproduise complètement le corps humain et sur lequel on pourrait faire porter les stimuli nécessaires à l'éducation[355]. Dans le cas des êtres humains, ces stimuli prennent la forme "d'interférences". Comme le terme l'indique lui-même, il peut être interprété positivement ou négativement et c'est d'ailleurs pourquoi Turing accorde à ces interférences un statut paradoxal : il reconnaît en effet qu'elles sont nécessaires mais qu'elles doivent aussi être évitées. Dans le chapitre 7 intitulé "Éducation de la Machinerie", il fait remarquer [356]:

> **«Il serait tout à fait injuste d'attendre d'une machine qui sort de l'usine qu'elle soit en mesure de se mesurer à un étudiant de doctorat. L'étudiant a eu des contacts avec des êtres humains depuis 20 ans ou plus. Ce contact a, pendant cette période, modifié sa démarche comportementale. Ses professeurs ont cherché de façon intentionnelle à la modifier. Nous pouvons dire que dans la mesure où l'homme est une machine, il est quelqu'un de sujet à beaucoup d'interférence. En fait, l'interférence semblera être la règle plutôt que l'exception. Il est en communication fréquente avec d'autres hommes et il reçoit continuellement des stimuli visuels et autres qui en eux-mêmes constituent une forme d'interférence. C'est seulement quand l'homme "se concentre" avec dans l'idée d'éliminer ces stimuli ou "distractions" qu'il approche de l'état de la machine sans interférence».**

Ces interférences sont par ailleurs considérées comme nécessaires dans le cas humain [357]:

> **« […] bien qu'un homme lorsqu'il se concentre puisse se conduire comme une machine sans interférence, son comportement quand il se concentre est largement déterminé par**

---

[355] Cf. **A. M. Turing**, "Intelligent Machinery", op. cit., p. 39 : « Une façon de se mettre à la tâche consistant à construire une "machine pensante" serait de prendre un homme dans sa totalité et d'essayer de remplacer toutes ses parties par des machines. (…) Ce serait évidemment une tâche énorme. (…). Pour que la machine puisse avoir une chance de découvrir des choses par elles-mêmes, on devrait l'autoriser à vagabonder dans la campagne et le danger que cela représente pour le citoyen ordinaire serait sérieux. De plus, même si les facilités dont on vient de parler lui étaient fournies, la créature en question n'aurait toujours pas de contact avec la nourriture, le sexe, le sport et beaucoup d'autres choses qui présentent un intérêt pour l'être humain. Aussi, bien que cette méthode soit probablement le moyen "sûr" de produire une machine pensante, il semble que cela soit finalement trop lent et peu praticable».

[356] **A. M. Turing**, "Intelligent Machinery", op. cit., p. 40.

[357] **A. M. Turing**, "Intelligent Machinery", op. cit., pp. 40-41. On ne peut pas ne pas penser à l'individu Turing en lisant ces lignes : qui d'autre que lui, pendant ses années de doctorat, a su s'identifier à une machine universelle en réduisant les "distractions" qu'il aurait pu avoir au "contact d'autres hommes" ?

la façon dont il a été conditionné par des interférences préalables».

Les interférences ou "distractions" possèdent donc elles aussi un statut ambigu, comme la peau "terminale" de l'esprit ou le "résidu de résidu" présent dans les machines. Les interférences proviennent à la fois de l'extérieur et sont de ce fait secondaires mais elles sont nécessaires à l'élaboration d'une intériorité chez les hommes. Ces interférences sont-elles ou non nécessaires dans le cas de la machine ? La réponse de Turing est tout aussi ambiguë : une machine universelle fonctionne sans interférence mais les interférences sont indispensables à l'apprentissage des machines particulières pour qu'elles se transforment en machines universelles. Dès lors, l'éducation des machines particulières, celles que "Computing Machinery and Intelligence" appelle les "machines-enfants", doit se faire de façon à imiter l'éducation humaine[358].

L'éducation humaine utilise comme interférence une alternance de récompenses et de punitions pour mener à bien l'apprentissage des enfants. Dans le cas des machines, on doit limiter les interférences à ce qui permet une communication *désincarnée* (que Turing appelle "non-émotionnelle") puisque les machines sont dépourvues de corps[359]. La communication doit donc être fondée sur un système de punitions et de récompenses différents dans le cas des enfants et dans le cas des machines [360] :

> **«L'utilisation des punitions et des récompenses peut, au mieux, faire partie du processus d'éducation. En gros, si le maître n'a pas d'autres moyens pour communiquer avec l'élève, la quantité d'information qui peut l'atteindre ne peut pas dépasser le total des punitions et des récompenses données. D'ici à ce qu'un enfant ait appris "Casabianca", il serait probablement très endolori s'il ne pouvait découvrir le texte que par la méthode des "vingt questions", où chaque "NON" impliquerait de recevoir un coup. Il est donc nécessaire d'avoir d'autres canaux "non-émotionnels" pour communiquer. S'ils sont disponibles, il devient possible d'enseigner par un système de punitions et de récompense à obéir aux ordres donnés dans un langage non-émotionnel, par exemple un langage symbolique».**

Dans le cas de la communication humaine qui doit s'instaurer entre le maître et

---

[358] **A. M. Turing**, "Intelligent Machinery", op. cit., p. 41 : «Ainsi, en appliquant une interférence adéquate, nous devrions espérer modifier la machine jusqu'à ce qu'on puisse compter sur le fait qu'elle réagisse par des réactions déterminées à certaines commandes».

[359] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 456.

[360] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 457.

l'enfant, Turing utilise l'exemple du poème "Casabianca". Il y a un étrange renversement dans la phrase où il est question du poème : en effet, il s'agit au départ d'un élève possédant un corps de chair et de sang qui doit *apprendre* le poème; mais cet apprentissage devient à la fin de la phrase la *découverte* du texte du poème lui-même, par la méthode des "vingt questions". On ne peut pas retrouver le poème "Casabianca" par induction, c'est-à-dire en se demandant, comme dans le dialogue au jeu de l'imitation décrit par Turing dans l'objection de la conscience [361], quel est le mot qui devrait suivre le mot précédemment découvert [362]. Dans le cas du poème "Casabianca", c'est la transformation de l'apprentissage en découverte, c'est-à-dire en *invention*, qui ne peut avoir lieu par le biais du canal de communication qu'est le corps et qui exige l'utilisation d'un canal désincarné, dans lequel la communication s'établit par le biais d'un langage de nature symbolique. Or ce langage est précisément celui de la machine. La transformation de l'apprentissage en invention passe donc par une *double transformation* : celle de l'être humain en machine et celle de la communication émotionnelle en une communication non-émotionnelle. On passe donc du cas humain au cas de la machine par cette double transformation.

La référence que Turing fait ici au poème "Casabianca" est donc complexe : elle renvoie à la fois au fait que la machine universelle, une fois existante, n'est plus soumise au type de processus inductif qui a permis sa création et au fait que c'est cependant par ce biais que Turing a lui-même réussi à accomplir son propre

---

[361] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 446.

[362] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 446. Il faut noter dans ce dialogue le fait qu'il ne s'agit pas d'un dialogue du jeu de l'imitation mais d'une parodie d'examen oral. Cette parodie ne rend cependant pas compte du fait que dans un examen oral, il n'est pas habituellement question, comme dans "Computing Machinery and Intelligence", d'"interrogateur" et encore moins de "témoin", ce qui donne à cette examen l'aspect d'un interrogatoire policier. Turing, deux ans après "Computing Machinery and Intelligence", à une époque où il commençait lui-même à faire l'objet d'une enquête policière, donna aussi l'apparence d'un interrogatoire policier à un dialogue du jeu de l'imitation : lors d'un débat contradictoire radiophonique qui se déroula à Manchester le 10 janvier 1952 sur la possibilité d'une intelligence mécanique, une personne participant au débat demanda à Turing : «Les questions devraient-elles être des additions ou pourrais-je lui demander ce qu'elle a eu au petit déjeuner ?» Turing : «Oh, oui, n'importe quoi, et les questions n'ont pas à être des questions, pas plus que dans un tribunal, les questions ne sont vraiment des questions. Vous voyez ce que je veux dire : "Je crois que vous faites seulement semblant d'être un homme" serait tout à fait approprié.» Cité dans **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 450.

auto-engendrement. Il y a bien là une “expérience” qui échappe à ce qui devrait faire partie du cursus normal de l’éducation d’un enfant : cette expérience est le passage de l’état humain à l’état mécanique. Or c’est précisément cette expérience qui permet d’introduire le mode de communication abstrait qu’est le langage symbolique : c’est donc cette expérience, aussi mystérieuse qu’elle paraisse, qui se situe à l’origine du langage symbolique. On peut donc dire que, de même que dans *Natural Wonders,* la question de la provenance de la première cellule n’était pas élucidée, de même ici la question de l’origine de la symbolicité du langage reste partiellement indéterminée. Ce que l’on rencontre quand on tente de préciser la nature du “résidu de résidu” ou de la peau “terminale” est la question de l’origine du langage. La question de l’origine du langage fait donc partie de l’aporie que nous avons appelée le “problème du continu” qui se situe au cœur du modèle computationnel de l’esprit.

Il est cependant possible de préciser l’origine de la symbolicité du langage à partir de la notion de peau, comme on va le voir maintenant.

5. Genèse de la peau

Telle qu’on l’a abordée jusqu’à présent, la notion de peau comme interface entre le corps et l’esprit est apparue comme une entité primitive à partir de laquelle il était possible de penser la machine. Il doit cependant être possible d’expliciter la constitution de cette notion. On doit alors retracer la genèse du rapport entre peau matérielle et peau psychique.

La réflexion de Turing sur la notion de peau, dans “Computing Machinery and Intelligence”, s’était constituée en deux étapes. La notion de machine avait tout d’abord été identifiée à l’esprit et c’est grâce à la notion de peau sans contenu qu’il a été possible de mettre en rapport les deux entités. C’est cette peau assimilée à une pure surface qui permettait ensuite à Turing de caractériser l’incarnation matérielle de l’identification entre l’esprit et la machine par deux traits au premier abord contradictoires, consistant à être à la fois mécanique *et* non-discrète. Le cheminement de la pensée de Turing conduit donc à concevoir une peau sans contenu, à la fois mécanique et non-discrète. La machine-*o* n’a donc pas d'autre statut que sa caractérisation psychologique.

On peut caractériser la machine par ces deux traits en apparence contradictoires parce que l'on doit distinguer en elle un contenant non-discret d'un contenu qui relève du signe discret, distinction qui vient se superposer à la différence entre l'intérieur et l'extérieur propre à la peau.

51. Les personnages du jeu de l'imitation

L'auto-création consiste donc à tenter de s'engendrer par le biais du concept de machine assimilée à une peau et malgré l'interdit sur la création bisexuée. Cette auto-création s'exprime par le biais du jeu de l'imitation dans lequel les données scientifiques et les données personnelles sont imbriquées les unes dans les autres.

On peut maintenant tenter de reconstituer les personnes à qui pensait Turing quand il a construit les règles du jeu de l'imitation, qui pourrait bien s'appeler, dans le roman intérieur de Turing, le jeu de la création. La femme qui est capable de créer par le biais de l'intérieur de sa peau, est sans doute Joan Clarke : de par cette création intérieure, elle est rapidement disqualifiée dans le jeu n°1. L'homme serait Turing lui-même et, par le biais de l'identification à la machine, deviendrait la machine universelle du jeu n°2. Quant à l'interrogateur, il pourrait bien s'agir de Christopher Morcom dont la perspicacité, aussi bien que l'influence *post-mortem*, serait prise désormais en défaut. Turing, grâce à sa position de survol dans le jeu n°2, aurait donc accompli son vœu le plus cher : réussir à assumer la vocation scientifique de Christopher Morcom par fidélité, comme le fait l'enfant de Casabianca à l'égard de son père.

52. La chimie de l'esprit

L'élucidation de l'énigme de la différence des sexes paraît dès lors à jamais inachevée et son expression par le biais du jeu de l'imitation, dans la mesure même où elle repose sur un équilibre entre des données personnelles, semble toujours instable.

Aussi l'auto-création peut-elle se transformer en *suicide* si le fragile équilibre inventé par Turing dans le jeu de l'imitation est remis en question. La condamnation que Turing encourut pour le délit d'homosexualité consistait à recevoir des injections d'hormones mâles, ces injections ayant pour but d'éradiquer son homosexualité en modifiant son équilibre chimique interne : il reçut les injections de février à mars 1953 durant une période dite de "probation". Ces injections eurent pour effet de le rendre temporairement impuissant et de lui faire pousser les seins[363]. Il ne recouvra sa pleine liberté qu'en avril 1953[364]. Un an plus tard, Turing se suicida, le 7 juin 1954 dans la soirée, en ingérant une pomme qu'il avait fait auparavant macérer dans du cyanure. Les circonstances de ce suicide, et en particulier, l'utilisation profondément symbolique d'une pomme, renvoient, là encore, à des faits de nature biographique. Outre le fait que Turing avait l'habitude de manger une pomme chaque soir avant de se coucher[365], Hodges rapporte, à propos de ce suicide, un certain nombre d'éléments.

Il remarque qu'il y a une certaine parenté entre le suicide de Turing et le rôle joué par Giacomo Casabianca dans le poème [366]: dans les deux cas, il s'agit d'un suicide qui prend la forme d'un sacrifice. Mais Hodges consacre surtout plusieurs remarques au rôle joué par la chimie dans l'esprit de Turing.

Premièrement, il rapporte qu'un film avait marqué Turing en Octobre 1937, "Blanche Neige et les sept nains" et tout particulièrement une scène, celle dans laquelle la marâtre de Blanche Neige tente de la tuer en lui faisant manger une pomme qu'elle avait fait auparavant macérer dans un bain de poison. La rime chantée par la marâtre, qui a les traits d'une sorcière, est celle-ci :

«Plonge la pomme dans le bouillon
Que la Mort qui endort s'y infiltre»

Hodges rapporte à ce propos que Turing avait l'habitude de chanter cette rime à n'en plus finir[367]. Hodges rapporte expressément le suicide de Turing à la scène du film[368] mais sans faire l'analyse de l'aspect symbolique de cet acte. Quand on sait que Turing n'a pas été élevé par sa mère mais par une nourrice, on

---

[363] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., pp. 473-474.
[364] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 486.
[365] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 279.
[366] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 515 : «Aucun des amis de Turing ne s'aperçut qu'il pouvait s'agir de quelque chose qui pouvait aider à rendre compte de sa mort, ni ne s'aperçut qu'il jouait finalement le rôle de Casabianca».
[367] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 149 et 489. Il y a, peut-être, une homophonie entre "Brewster" l'auteur de *Natural Wonders* et le bouillon [*brew*] de la chanson du film de Blanche-Neige.
[368] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 489.

peut se demander si la sorcière de Blanche-Neige et les sept nains n'est pas la nourrice en question ou même peut-être sa mère véritable. D'autres faits biographiques semblent aller dans le même sens[369].

Premièrement, Turing avait commencé en Octobre 1952 une psychanalyse avec le psychanalyste Franz Greenbaum. Greenbaum était un réfugié juif allemand et était de l'école jungienne. Hodges rapporte que Turing avait été frappé au cours de sa cure de découvrir l'hostilité qu'il avait à l'égard de sa mère[370].

Deuxièmement, Hodges rappelle aussi que son maître d'école écrivit un rapport aux parents de Turing sur sa conduite à l'école, rapport qui mentionne le fait qu'il avait surpris Turing en train de faire des expériences de chimie en chauffant un récipient à l'aide de deux bougies. Le maître d'école compare ces expériences à un «breuvage de sorcières»[371]. Hodges fait remarquer que c'est dans *Natural Wonders* que Turing commença à étudier la chimie qui y était seulement mentionnée par le biais de la description de poisons[372]. L'auteur de *Natural Wonders* comparait aussi le corps à une «boutique vivante d'apothicaire» et parlait de la découverte récente des hormones comme de messages chimiques que s'échangeaient les différentes parties du corps sans passer par le système nerveux [373]. Remarquons que Turing attachait, à la fin de sa vie, une signification très particulière au terme de poison : dans ses derniers articles portant sur la chimie de la morphogenèse, il utilise le terme de “poison” pour désigner le composant chimique qui empêche la croissance[374].

Troisièmement, un des amis de Turing, James Atkins, rapporte qu'après son article de 1936, Turing lui avait envoyé une lettre dans laquelle il se disait déprimé et considérait très froidement la possibilité de se suicider. Il allait même

---

[369] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 480.
[370] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 481.
[371] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 31, non traduit dans la trad. franç : «Je me moque de le trouver en train de faire bouillir Dieu sait quel brevage de sorcière grâce à deux chandelles coulant sur le rebord d'une fenêtre». L'allusion au brevage de sorcière fait sans doute allusion aux sorcières de *Macbeth*.
[372] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 17.
[373] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 26.
[374] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 494.

jusqu'à décrire ce suicide qui impliquait l'utilisation d'une pomme et d'un câble électrique [375]. On sait d'autre part, comme le rapporte Hodges[376], que Turing lut à cette époque le roman d'Angus Wilson, intitulé *Hemlock and after*, "La Ciguë et au-delà", publié en 1952.

Quatrièmement, il faut noter que l'ingestion d'un poison pouvait rappeler à Turing la mort de Christopher Morcom, son camarade de classe, dont le décès prématuré du fait d'une tuberculose bovine avait été causé par l'ingestion d'un lait empoisonné. C'est avec Christopher Morcom que Turing étudia la chimie, Morcom étant peut -être intéressé pour des raisons personnelles par ce sujet, lui qui savait que c'était du lait empoisonné qui l'avait rendu tuberculeux. Après la mort de Christopher Morcom, la famille Morcom institua un prix scientifique à l'école qui devait être remis chaque année à un élève : Turing reçut le prix la première année pour des recherches sur les équations mathématiques rendant compte de certaines transformations chimiques [377]. Le premier acte de fidélité à l'égard de la mémoire de Christopher Morcom passe donc par une réflexion mathématique sur des problèmes de chimie. Morcom semble donc être à la fois la première victime d'un "bouillon de sorcière" semblable à celui du film et celui qui montre comment écarter la menace de l'empoisonnement en faisant des poisons un objet d'étude mathématique. Le poison n'est objet d'étude que s'il se situe à l'extérieur du corps : s'il est transporté à l'intérieur du corps, c'est la volonté de tuer de la sorcière qui prend le dessus. Ce passage de l'extérieur à l'intérieur a donc directement rapport au domaine du chimique et du sexuel. En voici un autre exemple. En 1944, quand la nourriture était encore rare du fait du rationnement, Turing s'était procuré un livre sur les champignons qui venait d'être publié : il en faisait la cueillette et les rapportait à sa gouvernante pour qu'elle les prépare. Il aimait particulièrement répéter le nom du champignon le plus vénéneux, appelé en anglais "casquette de la mort" [*Death Cap*] ou Amanite

---

[375] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 129.
[376] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 487.
[377] **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 51.

phalloïde, qu'il chercha, sans succès, à trouver[378]. On retrouve ici la liaison entre le poison et la mort par l'intermédiaire du terme à connotation directement sexuelle de "phalloïde". La condamnation de Turing à recevoir une injection d'hormones mâles relève alors du même type de procédé : il s'agit de l'intrusion menaçante dans le corps d'un élément chimique ayant une signification sexuelle de par le transfert de l'extérieur à l'intérieur du corps. La condamnation légale de Turing entre peut-être ici en résonance avec la condamnation de son propre rêve d'engendrement. Turing disait d'ailleurs dans "Computing Machinery and Intelligence" :à propos de l'équivalence entre la construction des machines et la procréation des enfants[379]:

> **«Un argument d'une forme exactement semblable peut être fait dans le cas des machines. Cela peut paraître différent parce que c'est plus difficile à "avaler"».**

Aussi l'ingestion de la pomme empoisonnée, en réaction à ces injections d'hormone, a-t-elle aussi un autre aspect symbolique, pour qui a été élevé dans la tradition biblique : c'est en effet parce qu'Adam a voulu se substituer à Dieu en mangeant la pomme de la connaissance du bien et du mal qu'il est chassé du paradis terrestre.

Turing avait d'autre part imaginé en 1952 une espèce de récit de science-fiction et l'avait raconté à l'homme pour qui il allait être condamné. La trame du récit était la suivante : le hangar de la Royal Air Force à côté duquel Turing habitait était en fait un gigantesque cerveau-machine, programmé pour travailler pour tous ceux qui le désiraient. Mais quand il avait voulu s'en servir, il avait dû entrer dans le hangar et en était devenu prisonnier. Il devait, pour se libérer, jouer une partie d'échecs contre ce cerveau-machine. La machine jouait si vite qu'il fallait lui faire la conversation pour tâcher de la distraire de ses calculs. La machine ressentait alors de la colère. Puis il réussissait à battre la machine en faisant si mal de l'arithmétique qu'elle finissait par se suicider de désespoir[380].

---

378 **A. M. Turing**,"Computing Machinery and Intelligence", op. cit., p. 279.

379 **A. M. Turing**,"Computing Machinery and Intelligence", op. cit., p. 443.

380 **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 452. On remarque que, de même que la règle particulière qu'il avait inventée pour le jeu d'échecs et qui consistait à courir

Mentionnons enfin que Turing passa ses dernières vacances en compagnie de son psychanalyste et de sa famille[381]. Ils étaient allés passer le dimanche à Blackpool, près de Manchester, lieu d'une fête foraine permanente. Là, Turing était allé consulter une diseuse de bonne aventure, la "Reine Gitane", alors que les Greenbaum l'attendaient à l'extérieur : la consultation dura une demi-heure et Turing en ressortit livide et incapable de parler. Il les quitta le lendemain et les appela au téléphone le samedi suivant, deux jours avant sa mort, mais ils étaient absents. Turing s'est suicidé le lundi de Pentecôte 1954 - il avait quarante-deux ans -. Rappelons que le lundi de Pentecôte s'appelle en anglais [*Whit Monday*]. Le terme moyen-anglais de [*Whit*] a donné [*white*] ["blanc"] en anglais moderne : la coutume voulait que l'on baptise le dimanche de Pentecôte et que les nouveaux baptisés portent une robe de couleur blanche. On sait que, d'un point de vue théologique, la fête de Pentecôte en elle-même commémore la descente de l'Esprit-Saint sur les apôtres, le terme de [*wit*], d'une autre origine étymologique que [*whit*] mais homonyme, désignant l'esprit et le savoir de Dieu. On sait aussi que, d'un point de vue personnel, la Pentecôte était liée pour Turing au souvenir de Christopher Morcom, dont l'hymne préféré avait précisément pour sujet la Pentecôte[382]. C'est donc ce "lundi de l'esprit", dans lequel se mêlait à la fois une fête religieuse et le souvenir de son ami disparu qui fut fatal à Turing.

6. Remarques finales

Quelles conclusions tirer de cette analyse des réflexions de Turing sur la possibilité d'une "intelligence artificielle" à partir du jeu de l'imitation ?

---

entre les coups pour empêcher toute concentration, c'est le fait de distraire la machine de ses calculs qui provoque son désespoir. Cette "distraction" dont Turing parle aussi dans "Computing Machinery and Intelligence", est liée l'apparition du sexuel interprété comme ce qui empêche de calculer.

381 La relation avec son psychanalyste semble avoir évolué vers le registre de l'amitié pure et simple.

382 Cf. **A. Hodges**, *Alan Turing, The Enigma of Intelligence*, op. cit., p. 76. Voici l'hymne en question :

«*Gracious Spirit, Holy Ghost*
Taught by Thee we covet most
Of thy gifts at Pentecost
Holy heavenly Love »

La première remarque est que le jeu de l'imitation n'atteint pas le but pour lequel il est généralement invoqué : servir d'expérience de pensée montrant la légitimité du déplacement de la notion d'intelligence à d'autres substrats physiques que celui de l'espèce humaine. De ce point de vue, la preuve - ou tout au moins la confirmation - qu'était censé apporter le jeu de l'imitation est un échec. Il semble donc finalement encore plus difficile d'introduire un concept logique comme celui de machine de Turing en psychologie - comme tente de le faire Turing dans "Computing Machinery and Intelligence" - que d'introduire une réflexion psychologique sur le calcul dans le domaine logique - comme il parvient à le faire dans "On Computable Numbers …" -.

En revanche, le jeu de l'imitation décrit les conditions psychologiques d'invention du concept de machine de Turing telles qu'elles ont dû faire symboliquement sens pour l'individu Turing. Or ceci est un enseignement capital pour qui veut constituer à partir de la notion de machine de Turing un modèle de l'esprit. On peut invoquer deux raisons pour justifier cette remarque.

Premièrement, si la notion de machine de Turing peut servir de modèle à un certains nombres de processus de pensée et donc réaliser une modélisation au moins partielle de l'esprit - et on ne voit pas pourquoi ce ne serait pas possible -, elle laisse cependant un point aveugle dans cette modélisation, à savoir la question de l'invention du modèle lui-même et au premier chef, l'invention du concept de machine de Turing. Il y aurait une attitude singulièrement désinvolte et insidieusement positiviste à se désintéresser de cette question en croyant qu'elle relève de la pure contingence d'un destin individuel. Il faut réussir à comprendre les conditions d'invention de la notion de machine de Turing ainsi que son application au domaine de l'esprit parce que tout modèle de la psychologie se doit de rendre compte de l'originalité dont fait preuve l'esprit dans ses inventions ou ses découvertes. Or ceci ne peut pas être relégué à une pure affaire de contingence : il y a des *conditions* psychologiques de nature symbolique que nous nous sommes efforcés de décrire qui ont favorisé cet accès au concept de machine de

Turing[383]. Ces conditions ne sont pas descriptibles *a priori* et comme telles, elles n'entrent ni dans des catégories constituées à l'avance ni dans un réseau de causalité. Mais ce n'est pas une raison pour ne pas essayer de les décrire et d'en tirer profit pour essayer de rendre compte au mieux de la question psychologique de l'invention. Ces conditions relèvent plutôt du domaine *praxique* et c'est précisément leur appartenance à ce domaine qui pose le problème de la nature véritable de la notion de *représentation.* On avait vu qu'il était nécessaire de faire usage de deux sortes de symboles pour rendre compte de dans la notion de représentation[384]: les symboles cognitifs destinés à représenter des états du monde (par exemple les signes linguistiques ou les symboles chimiques) s'opposaient aux symboles praxiques destinés à donner sens à des actes (par exemple le signe de la croix ou les symboles des rêves et des jeux). Mais ce que l'analyse du "cas Turing" par le biais du jeu de l'imitation nous a appris, c'est que *cette opposition ne pouvait pas être une opposition radicale* : il y a à l'évidence des aspects praxiques dans la démarche de Turing, aussi "cognitive" qu'elle puisse être puisqu'elle se présente comme la théorisation du cognitif. C'est précisément l'intérêt de la notion d'invention que d'opérer de façon absolument unique une synthèse entre les aspects cognitifs et praxiques des symboles : l'invention de la notion de machine de Turing n'échappe pas à cette règle. On peut, certes, essayer de constituer à partir de la notion de machine de Turing un modèle "tout cognitif" de la psychologie. Mais la rançon nous parait être lourde à payer du point de vue du modèle ainsi constitué : comment croire à la pertinence du modèle en question si l'on élimine d'office des manifestations psychologiques aussi typiquement "humaines" que, par exemple, les symboles des rêves et des jeux ? C'est la nature du rapport entre les deux types de symboles qui doit permettre de préciser ce que l'on entend par représentation.

Deuxièmement, le modèle de la psychologie que l'on peut fonder à partir

---

[383] Il est à noter que ces conditions n'ont rien de "rationnelles" au sens étroit du mot. Leur description tend à montrer, s'il en était besoin, que l'histoire des sciences décrite comme histoire d'une rationalité purement abstraite est bien loin de la vérité. L'histoire des sciences fait partie de l'histoire tout court et n'a pas de privilège rationnel particulier du point de vue des conditions psychologiques qui président à sa constitution.

[384] Cf. Introduction , deuxième partie, § 21.

du concept de machine de Turing doit pouvoir montrer comment il est psychologiquement possible de s'élever à l'universalité dans l'invention d'un concept. C'est précisément dans cette "induction", pour reprendre le terme de Turing, que les symboles praxiques jouent un rôle. Le modèle de la psychologie que l'on peut constituer à partir de la notion de machine de Turing ne doit donc pas consister à plaquer une structure calculatoire - fût-elle universelle - sur la notion d'esprit réduite dès lors au statut d'objet mais il doit pouvoir montrer comment on peut graduellement faire surgir une structure calculatoire d'un objet qui ne l'est pas au départ. C'est dans cette gradation qu'il doit être possible d'articuler les aspects cognitifs et praxiques des symboles.

Cette démarche est fidèle à celle de Turing dans le jeu de l'imitation puisque, comme nous l'avons montré, c'est exactement le sens de son argumentation probabiliste. C'est par ce biais qu'il devient possible de généraliser le point de vue de Turing et d'en tirer profit pour une conception générale de l'esprit. En effet, on sait que les conditions psychologiques qui ont permis à l'individu Turing d'inventer le concept de machine de Turing ne peuvent pas valoir pour un autre individu : on a vu qu'elles lui étaient éminemment personnelles. Mais elles sont pourtant généralisables : elles montrent d'une part qu'il est nécessaire *qu'il y en ait* pour que les individus puissent parvenir à inventer et d'autre part qu'elles articulent toujours le physique *et* le mental, même quand elles apparaissent comme privilégiant un terme sur l'autre. Le cas personnel de Turing portait sur l'analogie entre la différence des sexes et la différence entre le continu et le discret : la façon dont il concevait cette analogie lui était éminemment personnelle mais les deux différences qui la composent lui préexistent et demandent à être remise en chantier par chacun. C'est précisément cette dimension de travail *personnel* que doit prendre en compte un modèle général de l'esprit.

Il faut donc étudier maintenant comment on peut articuler les deux grandes oppositions que l'on vient de décrire, celle qui porte sur l'aspect cognitif et praxique des symboles et celle qui porte sur l'aspect individuel ou général des conditions psychologiques qui ont présidé à l'invention d'un concept.

Troisième partie

La portée générale du modèle de Turing

«Par l'intermédiaire d'une certaine interprétation, les énoncés revêtus d'une validité axiomatique auront à assumer une seconde signification et à subir une seconde épreuve de validité - cette fois en tant que lois naturelles régissant un certain ensemble de phénomènes».

F. Gonseth, *Le problème du temps*, p. 336.

Introduction

Quand on demande quelle portée pourrait avoir le modèle de la psychologie tel qu'il a été élaboré par Turing, on se heurte immédiatement à une difficulté qui touche à la généralité du modèle en question. On a vu en effet que l'élaboration du projet d'une intelligence artificielle avait non seulement été le fruit de circonstances historiques particulières comme la constitution progressive d'une science et d'une technique informatique après la seconde guerre mondiale mais qu'elle reposait aussi en partie sur des interprétations les plus étranges quand on analysait ce qui en constituait l'un de ses fondements, l'analogie établie par Turing entre le discret et le continu d'une part et l'homme et la femme de l'autre. Que pourrait-on alors espérer en déduire d'un point de vue général, c'est-à-dire du point de vue de la caractérisation proprement conceptuelle de la notion psychologique d'intelligence ?

A première vue, l'analogie que Turing constitue lui est trop personnelle pour que l'on puisse espérer pouvoir en tirer des conclusions sur la portée générale du modèle. Nous allons cependant tenter d'émettre un certain nombre d'hypothèses touchant cette généralisation et ce, dans deux directions, celle de la notion de représentation d'une part et celle de l'analyse de la notion d'invention d'autre part.

1. La notion de représentation

La question du statut épistémologique du modèle de Turing implique, comme on va le voir à présent, de s'interroger sur le statut de la notion de représentation dans la constitution du projet de l'intelligence artificielle. Que faut-il entendre par le terme de “représentation” ? Reprenons la description qu'en donne C. P. Bruter [385]:

> **«Nous avons noté que l'activité première de notre cerveau est de réguler notre milieu intérieur et de simuler le monde extérieur. Or, qui dit simulation d'un phénomène dit réplique des formes des objets, et des trajectoires qu'ils suivent au cours du procès. L'activité principale de notre cerveau, en second après sa fonction régulatrice, est donc de fabriquer des modèles de description de notre univers».**

Entre la fonction régulatrice de l'organisme et sa fonction simulatrice, le modèle psychologique de Turing a cependant montré l'existence d'une troisième fonction : la fonction *dissimulatrice*.

C'est en effet tout le sens du modèle de Turing de montrer que la simulation n'est qu'une partie d'une notion plus large qui est celle d'imitation. Or l'imitation peut signifier simulation comme le remarque Bruter et comme l'a montré Turing dans "On Computable Numbers …" mais elle signifie aussi dissimulation, comme Turing l'a montré dans "Systems of Logic based on Ordinals" en employant la notion de machine-*o* et comme il l'a définitivement établi dans "Computing Machinery and Intelligence" par le moyen du jeu de l'imitation. La fonction de dissimulation ne doit pas être interprétée de façon péjorative : elle permet de manifester ce qui, autrement, resterait caché. Le jeu de l'imitation est bien, de ce point de vue, un chef-d'œuvre de dissimulation : il permet à Turing de dévoiler la façon dont il réfléchit l'invention de ses propres concepts et ce, par le biais de trois images, celle de "l'équipe d'ingénieurs", celle de la "peau de l'oignon" et celle de la "machine-enfant", images qui lui permettent de constituer cette analogie si curieuse entre le discret et le continu d'une part et l'homme et la femme d'autre part. La fonction de dissimulation permet ainsi l'élaboration de contenants de pensée dans lesquels peuvent être reçus à titre de contenu les états internes de l'organisme et les états externes du monde. La dissimulation renvoie de ce fait à la notion de peau comme contenant primordial délimitant un intérieur d'un extérieur. La fonction de dissimulation rend ainsi possible la constitution d'un domaine virtuel au sein duquel l'usage d'images permet d'articuler le plan de l'organisme et de sa régulation et le plan du

---

[385] **C. P. Bruter**, *Sur la nature des mathématiques*, Gauthier-Villars, Paris, 1971, p. 3.

langage cognitif et sa faculté de simulation.

Dans cette mesure, une interprétation “toute cognitive” du modèle de Turing ne peut pas rendre pleinement compte de la nature de la représentation. Aussi est-il vain de considérer le modèle proposé par Turing comme un paradigme utilisable tel quel pour servir de fondement à une intelligence artificielle. Si l’on abandonne la recherche d’un fondement de ce type sans vouloir cependant abandonner la portée explicative du modèle proposé par Turing en psychologie, on se trouve alors confronté au problème de la place que l’on doit accorder à la notion de représentation au sein du modèle.

Nous tenterons de montrer que l’absence de reconnaissance du rôle médiateur de la fonction dissimulatrice a pour conséquence la constitution d’interprétations dualistes de la notion de représentation. Ce dualisme prend deux formes : une forme platonicienne et une forme nominaliste. Ces interprétations dualistes rendent impossible la compréhension du rapport entre l’aspect cognitif et praxique des symboles, contrairement aux interprétations non-dualistes de la représentation qui tentent de rendre compte du modèle à partir de son invention et que nous étudierons pour finir, en essayant de conférer une portée générale à l’analogie élaborée par Turing au fondement de son modèle.

2. La notion d’invention et le rôle du langage

L’élaboration du modèle de Turing permet en effet d’éclairer d’un jour nouveau la question si controversée de l’invention des concepts et en l’occurrence de l’invention d’un concept de nature logico-mathématique, celui de machine de Turing.

En fait, Turing n’a pas cherché à inventer un modèle scientifique pour la psychologie. Turing est un mathématicien avant tout et comme on l’a vu, c’est la question de la caractérisation des réels calculables qui l’a occupé au premier chef. Mais l’aspect inévitablement psychologique de la caractérisation du concept de fonction calculable l’a conduit, ultérieurement et contrairement aux mathématiciens qui ont proposé d’autres formalismes visant la caractérisation du même objet, à dévoiler ce qui lui semblaient être les conditions psychologiques qui ont présidé à l'invention du formalisme de la machine de Turing. Pourquoi la caractérisation des fonctions calculables par le biais du concept de machine de Turing a-t-elle eu cet effet ?

On sait l'importance en mathématique d'une "bonne" notation pour désigner un concept nouveau[386]. On peut supposer que la notion de machine fait office, dans le cas des fonctions calculables, de "bonne" notation, précisément parce que, d'un point de vue psychologique, elle est parlante pour l'intuition, alors que les notions de fonction récursive ou de fonction λ-définissable sont immédiatement plus abstraites et de ce fait, ne rendent pas directement compte de l'aspect psychologique de la caractérisation du concept de fonction calculable. Dans cette hypothèse, la "bonne" notation du point de vue psychologique serait non pas celle qui possède un aspect exclusivement cognitif - décrivant un état du monde - mais celle qui articule un aspect cognitif et un aspect praxique - donnant sens à un acte -. C'est Turing qui aurait ainsi découvert ce qui, pour la psychologie du mathématicien au travail, serait la "bonne" notation pour le concept de fonction calculable : la notion de machine de Turing. Cette notation permettrait d'articuler l'aspect mathématique et l'aspect psychologique présents dans le concept. Plus précisément, c'est parce que la notion de machine peut à la fois être décrite par les signes discrets du langage (c'est-à-dire par des symboles cognitifs) et par des images qui renvoient à la peau et donc à la nature corporelle du sujet (c'est-à-dire par des symboles praxiques) que la notion de machine parvient à articuler le domaine mathématique et le domaine psychologique : c'est cette articulation qui

386 Leibniz en particulier a remarqué l'importance de la notation pour parvenir à une caractérisation adéquate des concepts : «[…] l'instrument général de l'invention humaine réside dans des caractères appropriés, ce que montrent suffisamment les exemples de l'arithmétique, de l'algèbre et de la géométrie : l'esprit se trouve de ce fait dirigé par un fil quasi sensible pour qu'il ne se perde pas comme dans un labyrinthe. Dans la mesure où l'esprit ne peut pas embrasser de nombreuses choses d'un même mouvement, son imagination est épargnée quand les signes remplacent les choses : il est très important que la façon dont les signes sont constitués puissent permettre de référer adéquatement aux choses. Ma propre contribution au progrès des mathématiques est née d'un seul fait : avoir rendu meilleur l'usage des symboles représentant les quantités». **Leibniz**, "Inventorium Mathematicum", *Mathematische Schriften*, ed. Gerhardt, Olms Verlag, Hildesheim, 1863, Band VII, p. 17. C'est ce que remarque aussi Cavaillès dans *Méthode axiomatique et formalisme; essai sur le fondement des mathématiques*, Hermann, Paris, 1938, réédition 1981 et plus particulièrement dans le § intitulé "Philosophie du signe", pp. 91-95 de la réédition.

rend possible la constitution d'un modèle pour la psychologie à partir du concept de machine de Turing qui, autrement, resterait un concept exclusivement logico-mathématique.

De ce point de vue, la possibilité d'un "sacrifice" de la peau rend possible l'élaboration d'images permettant à Turing de décrire l'articulation entre l'aspect mathématique et l'aspect psychologique de la notion de machine de Turing. Les trois images employées par Turing lui sont éminemment personnelles dans la mesure où elles lui permettent de dévoiler sa propre attitude à l'égard de la différence des sexes et à l'égard de la création en général. Mais l'usage d'images pour constituer un modèle de la psychologie n'est pas proprement individuel [387] et c'est donc cet usage qui porte en lui la possibilité d'une généralisation : le fait que la peau joue un rôle moteur dans la constitution des trois images souligne en effet la fonction *générale* de tout symbole praxique qui est de rapporter au corps le sens accordé à l'acte que le symbole praxique exprime.

L'expérience de pensée du jeu de l'imitation a donc permis de dévoiler la prégnance des symboles praxiques au sein d'un domaine qui semblait inaccessible à ce type de symboles. La présence de symboles praxiques dans un modèle réputé "tout cognitif" ne nous semble pas ruiner pas l'intérêt du modèle en question mais le situe au contraire à sa vraie place dans la constitution d'un modèle du domaine très spécifique de la *psychologie* qui requiert une méthodologie particulière. L'esprit est un objet très particulier dans la mesure où il ne peut pas être décrit comme un simple état du monde : il est cet objet du monde par les actes duquel existent des états du monde. C'est pourquoi il paraît nécessaire, pour constituer un modèle adéquat de la psychologie, de prendre en compte les deux types de symboles et leur articulation.

Cette remarque doit nous servir du point de vue de l'analyse générale de la notion d'invention : l'invention est un *acte qui donne sens* à des matériaux qui jusqu'alors n'avaient pas de signification les uns par rapport aux autres. C'est la découverte d'une *connexité* entre des éléments considérés auparavant comme épars

---

[387] Nous l'avions déjà noté dans l'avant-propos de notre travail au § 422 en mentionnant le titre d'un ouvrage de Michael Arbib, *The Metaphorical Brain; an Introduction to Cybernetics as Artificial Intelligence and Brain Theory*.

qui en est le fondement psychologique. Aussi n'est-il pas possible de rendre compte d'une invention en ne prenant en compte que l'aspect cognitif des symboles utilisés parce que les symboles cognitifs ne peuvent décrire que des états déjà existants du monde et non pas des états encore virtuels qui cherchent à s'incarner dans une structure organisée. C'est précisément à la description de cette structure qu'il faudra s'attacher.

Chapitre I

Interprétations dualistes
du modèle de Turing

Les interprétations dualistes du modèle de Turing prennent pour point de départ la distinction entre le corps - généralement limité au cerveau - et les manifestations de l'esprit et tentent à partir de cette distinction de fonder un modèle, généralement appelé “modèle computationnaliste de l'esprit”. Le modèle computationnaliste de l'esprit associe les états mentaux à des états de machine de Turing. Il accorde une place particulièrement importante à la notion de machine universelle parce que la possibilité d'une adaptation de l'esprit humain à des tâches multiples semble analogue à la capacité universelle de simulation de ce type de machine[388].

Les deux interprétations dualistes les plus courantes sont soit platonicienne soit nominaliste. Il est de ce point de vue remarquable de constater que le modèle computationnaliste de l'esprit *peut être justifié par des présupposés philosophiques radicalement opposés*, dès lors que l'on accepte la distinction entre le corps et l'esprit. Aussi faut-il voir dans les justifications philosophiques du modèle computationnel des justifications autant apportées à cette distinction première qu'au modèle lui-même.

Ce chapitre poursuit deux buts. D'une part, il vise à décrire les deux grands types de justification philosophique du modèle en question et à montrer qu'ils ont

---

[388] Cf. **J. Haugeland**, “Semantic Engines : an Introduction to Mind Design” in [**J. Haugeland ed.**, *Mind Design*, op. cit.], p. 13.

bien pour origine la distinction radicale opérée entre le corps et l'esprit. D'autre part, l'aspect presque arbitraire de la justification philosophique du modèle dualiste permettra de corroborer de façon indirecte la possibilité du point de vue non-dualiste qui est le nôtre, ainsi que la justification philosophique qui l'accompagne.

1. Les deux conceptions philosophiques à la racine du dualisme du modèle computationnel

Le dualisme computationnel est engendré par deux conceptions radicalement opposées de ce qu'il faut entendre par pensée : soit le modèle est justifié par une conception platonicienne pour laquelle penser, c'est raisonner sur des idées; soit le modèle est justifié par une conception nominaliste, pour laquelle penser, c'est raisonner sur des signes. Ces deux conceptions s'opposent sur la façon dont elles conçoivent le rapport de la pensée au langage.

Dans le cas platonicien, la pensée existe indépendamment du langage qui n'est qu'un moyen expédient d'assurer la communication entre les personnes : la genèse de la pensée se trouve ainsi dissociée du langage.

Dans le cas nominaliste, les signes du langage sont, pour chaque individu, le *seul* terrain qui permette d'avoir accès à sa pensée et à celle d'autrui : les deux domaines de la pensée et du langage sont à tel point coextensifs qu'ils s'identifient presque[389]. Dans le cas particulier du modèle computationnel de l'esprit, on note l'apparition d'une opposition interne à cette conception, opposition qui se situe entre les tenants de ce qu'il est convenu d'appeler le "fonctionnalisme" et les tenants du "physicalisme". Pour le fonctionnaliste, le domaine des états mentaux a une autonomie et il peut être décrit grâce au modèle de la machine de Turing. Pour le physicaliste, on doit pouvoir en droit réduire tous les états mentaux à des propriétés physico-chimiques du cerveau. Comme le remarque N. Block, il est curieux de constater que le fonctionnalisme sert tantôt à

---

[389] C'est le cas chez Ockam pour qui l'universel de la pensée est un *signe* mental et naturel, tandis que le mot est un signe concret et arbitraire.

défendre le physicalisme tantôt à le récuser [390]. En attribuant une certaine autonomie à un niveau logique susceptible de modéliser les états mentaux, il devient impossible de justifier le rapport de ce niveau à celui de la matière physique. Il faut y voir la marque d'un conflit dialectique[391] qui naît de la distinction entre le corps et l'esprit.

Les deux attitudes, platonicienne et nominaliste, si opposées qu'elles soient entre elles, n'en engendrent donc pas moins un dualisme identique quand elles sont appliquées au problème particulier de la constitution d'un modèle de l'esprit. La seule différence vient de ce que dans le cas platonicien, il est pleinement assumé et revendiqué, alors que dans le cas nominaliste, il est plutôt vécu sur le mode du "remords". Cela tend à prouver que dès que l'on cherche à établir un parallèle entre les états du cerveau et les états récursifs d'une machine universelle de Turing, *quel que soit le présupposé philosophique dont on part*, on est conduit à se justifier au moyen d'une théorie qui envisage l'esprit comme séparé de la matière; bref, que c'est bien le présupposé de la distinction du corps et de l'esprit qui gouverne les présupposés philosophiques en question. Étudions les formes prises par ce dualisme.

11. Le modèle platonicien de l'esprit : le cerveau et l'âme

Le grand représentant contemporain de l'attitude platonicienne est K. Gödel dont le platonisme[392] en mathématiques a engendré, pour ce qui est du

---

[390] «Certains philosophes considèrent que le fonctionnalisme montre que le physicalisme est probablement exact tandis que d'autres considèrent qu'il montre que le physicalisme est probablement inexact». **N. Block**, "What is Functionalism ?" in [**N. Block ed.**, *Readings in the Philosophy of Psychology*, I, Harvard University Press, 1980], p. 177.

[391] Le terme est pris dans son acception kantienne.

[392] Le platonisme au sens de la doctrine historique de Platon a au moins ceci de commun avec le platonisme en philosophie des mathématiques qu'il prend pour point de départ la constatation de l'existence d'un indécidable (qui se manifeste par des apories dans les définitions de la science, du langage, de la vertu, du courage, etc), la pensée consistant à prendre une *bonne décision* quant à cet indécidable, c'est-à-dire à prendre une décision conforme à la nature. Dans le platonisme contemporain en philosophie des mathématiques, la question de l'indécidable est ainsi liée à cet "optimisme" rationaliste qui fait éthiquement espérer que tout ce qui apparaît sous la forme aporétique de l'indécidable sera, ultérieurement, dépassé par une bonne décision. C'est précisément le cas de Gödel pour qui les limitations des formalismes telles qu'il a contribué à les déterminer ne sont qu'une étape qui sera ultérieurement dépassée par la découverte de nouveaux axiomes plus adéquats en théorie axiomatique des ensembles.

fondement d'une théorie de l'esprit, un modèle dualiste. Les implications dualistes de l'attitude platonicienne en théorie de l'esprit ont été analysées chez Gödel par Wang Hao. Celui-ci remarque [393]:

> **«Il ne me paraît pas évident que ce qui peut être observé dans le monde physique soit toujours récursif (ou "mécanique"). Par exemple, en 1971, je posai à Gödel la question de savoir si une machine physique pouvait être construite de telle sorte qu'elle puisse produire des suites non récursives, interprétées comme des conséquences de son caractère matériel. Gödel répondit par la négative, mais je ne pus comprendre les raisons qu'il donna pour atteindre cette conclusion. Une question moins vague est peut-être de demander si tout calculateur analogique peut être simulé par un ordinateur digital de façon complètement adéquate (compte tenu des probabilités définies d'erreur).**
>
> **De manière spécifique, Gödel exhibe deux propositions : (i) le cerveau fonctionne fondamentalement comme un ordinateur digital. (ii) les lois physiques, dans leurs conséquences observables, ont un degré fini de précision. Il pense que (i) est très probable et (ii) pratiquement certain (cf Wang Hao, *From mathematics to philosophy*, p. 326). Je ne vois pas de raison convaincante pour croire que (i) est vrai (sauf peut-être si l'on souhaite faire une place pour un esprit séparé). Je ne suis pas sûr de savoir ce que (ii) veut dire».**

Revenons un instant sur le raisonnement de Gödel tel qu'il est reproduit par Wang Hao. Gödel pensait que la question : "En quel sens l'ordinateur digital est-il un modèle de l'esprit humain ?" reposait, pour être intelligible, sur deux présupposés[394]. Premièrement, le cerveau fonctionne comme un ordinateur digital. Deuxièmement, les lois physiques, dans leurs conséquences observables, ont un degré fini de précision.

Gödel pensait que ces deux présupposés étaient liés, plus exactement que le deuxième était un énoncé affaibli du second : parce que les lois physiques ont un degré fini de précision, ce qui est observable d'un point de vue physique, c'est ce qui est récursivement calculable; aussi le cerveau en tant qu'objet physique fini peut-il être assimilé à un ordinateur digital parce que ses sorties, pour être causalement explicables, relèvent *par définition* de ce cadre.

Les deux présupposés sont-ils liés de la manière dont l'entendait Gödel ?

---

[393] **W. Hao**, *Reflections on Kurt Gödel*, MIT Press, Cambridge, Mass. 1987, p. 198 et aussi **J.-P. Dubucs**, "La philosophie de Kurt Gödel", L'âge de la science, Gallimard, 1991.

[394] Notons que Hao Wang présente la pensée de Gödel de façon légèrement différente dans son ouvrage plus ancien *From Mathematics to Philosophy*, op. cit., p. 326, puisque le premier présupposé que Gödel considère comme étant un «préjugé de l'époque» est qu'il n'y a pas d'esprit séparé de la matière. La nouvelle présentation a l'avantage de dissocier d'une part l'identification du cerveau à un ordinateur digital - que Gödel accepte - et d'autre part le monisme matérialiste qui ne conçoit pas l'existence d'un esprit séparé - que Gödel refuse -.

Comme Wang Hao le remarque dans son commentaire, ce “degré fini de précision” ne peut pas vouloir dire que toutes les lois physiques, pour être considérées comme physiques, doivent être mécaniques ou récursives, puisqu’on a trouvé des contre-exemples physiques à cette identification[395]. Ceux-ci ont montré qu’il y a possibilité d’isoler, au sein du monde physique, des phénomènes qui dépassent le cadre de la calculabilité : si l’on peut *observer* des résultats non-récursifs, c’est-à-dire des phénomènes dont la causalité physique nous échappe, alors on doit en conclure que causalité physique et description récursive ne se recouvrent pas[396]. Le “degré fini de précision” ne peut donc pas renvoyer à l’aspect calculable des lois de la physique.

Aussi faut-il, pour donner un sens à l’interprétation de Gödel, ajouter une condition restrictive, comme le fait Wang Hao, à l’expression de “degré fini de précision” : les lois physiques ont une *limite finie et uniforme* de précision[397]:

> **«Si nous considérons ces conditions initiales comme des entrées et la solution comme une sortie, nous aurions quelque chose qui ressemblerait à une machine physique produisant des sorties non récursives; la question de la construction d’une telle machine semble être une question pratique. Si Gödel veut dire par le présupposé (ii) que seules les lois mécaniques sont précises, alors le résultat que l’on vient de mentionner semble réfuter (ii). Il semble, quoi qu’il en soit, que (ii) est plus plausible que ce que cette interprétation suppose. Assurément, on serait enclin à penser que toutes les observations ont des limites finies de précision. Mais alors (ii) semble ne rien à voir avec (i), alors que Gödel envisage (ii) comme un affaiblissement de (i) qui préserve quelque chose de son contenu. Peut-être que Gödel veut dire qu’il y a une limite uniforme finie de précision à toutes les conséquences observables des lois physiques. Une telle interprétation de (ii) semble le rendre compatible avec la présence de lois physiques non mécaniques et en même temps le relier à (i), au sens où pour tout ce que l’on peut observer du fonctionnement du**

---

395 Hao Wang cite le résultat de Pour-El et Richards (**M.-B. Pour-El** et **I. Richards**, “A computable differential equation which posesses no computable solution”, Ann. Math. Log, 17, pp. 61-90) qui ont prouvé que l’on peut trouver des conditions initiales calculables pour une équation d’onde, conditions engendrant une solution unique qui, bien que continue, n’est pas calculable. Cette constatation forme le point de départ du travail de Penrose et de son équipe concernant la possibilité d’un modèle du cerveau qui ne soit pas limité au cadre strict de la récursivité. Ces recherches prennent à l’heure actuelle deux directions : l’une physique et l’autre biologique. Cf. **R. Penrose**, *The Emperor’s New Mind*, op. cit., p. 558 pour la direction physique et p. 516 pour la direction biologique. Penrose, dans une conférence récente (Oxford, 6-6-93), s’appuie aussi sur les travaux de Stuart et Hameroff qui étudient l’activité de contrôle moteur des paramécies, organismes monocellulaires qui n’ont pas de système nerveux et qui ne peuvent donc pas opérer de contrôle moteur par le biais d’un calcul.

396 Et ceci, contrairement à ce que l’on trouve quelque fois dans la littérature. Cf. par exemple, **W. Poundstone**, *The Recursive Universe*, Oxford University Press, Oxford, 1987, en particulier chap. 7 : “Physics as recursion”. La même idée se trouve également dans **J. Ramunni**, *La physique du calcul; Histoire de l’ordinateur*, op. cit., p. 216.

397 **W. Hao**, *Reflections on Kurt Gödel*, op. cit., p. 198.

cerveau, celui-ci peut être traité comme un ordinateur digital».

Cette restriction permet, pour Wang Hao, de conserver le rapport entre les deux présupposés que Gödel voulait associer[398]. En effet, en introduisant des considérations sur une limite uniforme et finie de précision, on reconnaît l'existence de quelque chose qui dépasse la limite : dès lors, on doit distinguer la causalité d'une part et la récursivité de l'autre.

C'est ce raisonnement, légèrement précisé par Wang Hao, qui a, comme celui-ci le remarque lui-même, des conséquences dualistes pour ce qui est de la constitution d'un modèle de l'esprit. En effet, dès que l'on s'accorde sur la nécessité de distinguer la causalité et la récursivité, s'offre la possibilité de concevoir un cerveau capable de *produire* causalement autre chose que ce qui est récursivement *descriptible*, sans que la causalité produisant cet "autre chose" puisse jamais être entièrement assignable à une loi récursive. Le raisonnement de Gödel autorise ainsi à penser que c'est précisément la récursivité des programmes capables d'être exécutés par une machine universelle de Turing qui implique d'accorder un fonctionnement non-physique, et partant non-matériel, au cerveau. Gödel était cohérent avec lui-même : parce qu'il considérait que le cerveau était descriptible comme une machine digitale, il admettait l'existence d'un esprit séparé de la matière[399]. Ce dualisme de l'esprit qui distingue dans cette notion un

---

398 La restriction de Wang Hao paraît être conforme avec l'attitude philosophique générale de Gödel : celui-ci disait que ses résultats d'incomplétude, théorèmes de la théorie finitaire des nombres, avait été obtenus parce qu'il s'était intuitivement placé du point de vue de l'existence du transfini. De cette attitude découle la nécessité de discerner la *limite finie et uniforme* de précision d'un énoncé (en l'occurrence de toute axiomatique) faisant référence à une réalité dont l'existence n'est pas récursivement accessible. Gödel avait d'ailleurs noté lui-même une "profonde analogie" entre la théorie des ensembles et la physique théorique. Cf. **W. Hao**, *Reflections on Kurt Gödel*, op. cit., p. 52

399 Pour Gödel, il y a deux résultats rigoureusement prouvés concernant les rapports de l'esprit et des machines. Premièrement, l'esprit humain n'est pas capable de mécaniser toutes ses intuitions mathématiques : quand il y parvient en constituant un formalisme, cette constitution elle-même sert de base à une nouvelle intuition, celle de la consistance de ce formalisme. Deuxièmement, on a la disjonction suivante : ou bien l'esprit humain peut décider plus de questions arithmétiques que n'importe quelle machine, ou bien il y a des questions arithmétiques indécidables pour l'esprit humain. Le platonisme de Gödel lui fait admettre, dans le premier résultat, l'hypothèse d'une intuition mathématique infinie en constant développement et le fait pencher, dans le second, vers la première branche de l'alternative car une indécidabilité *absolue* lui paraît irrationnelle : l'esprit humain se poserait des questions auxquelles il serait incapable de répondre. Ces deux résultats sont rapportés par Wang Hao dans **W. Hao**, *From mathematics to philosophy*, op. cit., p. 324.

cerveau et une âme exige que soit séparé d'une part le fonctionnement du cerveau, nécessairement lié au passage du temps et d'autre part l'existence désincarnée de l'âme qui n'est pas soumise à ce passage. L'abolition du temps, au moins comme *possibilité cosmologique*, a été par ailleurs décrite par Gödel comme s'intégrant au cadre de la relativité générale d'Einstein[400].

Pour Gödel, la théorie mécaniste de l'esprit n'implique donc *en aucune façon* d'adhérer à une conception moniste qui ferait de la notion d'intelligence un pur concept entièrement descriptible par le formalisme de la machine de Turing. Il s'oppose de ce point de vue à l'attitude nominaliste.

12. Le modèle nominaliste de l'esprit

La thèse nominaliste apparaît comme l'attitude dominante quand on cherche à justifier le modèle computationnel de l'esprit. Elle a donné lieu, en théorie de l'esprit et contrairement à l'attitude platonicienne, à de multiples recherches qui se subdivisent en courants philosophiques nombreux. Nous n'analyserons pas toutes les ramifications de ces courants et nous ferons seulement remarquer que l'attitude nominaliste engendre, elle aussi, une attitude dualiste à partir de la distinction du corps et de l'esprit.

On sait que le nominalisme est une doctrine concernant le statut du langage qui, à l'inverse de l'attitude platonicienne, refuse toute existence aux concepts universels extra-mentaux et ne voit en eux que des signes que l'on manipule selon des règles, en vue de désigner des objets matériels particuliers[401]. Aussi toutes les expressions qui, faisant usage de termes abstraits, leur confèrent une existence, comme par exemple les expressions qui enveloppent l'usage de l'infini, sont-elles

---

400 Il s'agit de trois articles publiés de 1949 à 1952 et republiés dans le tome II des *Collected Works*, Oxford University Press, Oxford, 1990. On trouvera des commentaires sur ces articles dans **J. Merleau-Ponty**, *Cosmologie du XXème siècle*, Gallimard, Paris, 1965, pp. 276-286 ainsi que dans **P. Yourgrau**, *The Disappearance of Time, Kurt Gödel and the Idealistic Tradition in Philosophy*, Cambridge University Press, Cambridge, 1991, particulièrement chap. 3 "Time travel and the Gödel universe". Gödel parvient ainsi à conclure qu'«[…] il devient théoriquement possible de voyager dans le passé ou de l'influencer en quelque manière». **K. Gödel**, "An example of a new type of cosmological solutions of Einstein's field equations of gravitation", Review of modern physics, 21, republié dans [**K. Gödel**, *Collected Works*, t. II, op. cit.], p. 191.

401 Comme le fait remarquer N. Goodman dans **N. Goodman**, *The Structure of Appearance*, Dordrecht-Boston, 1977.

condamnées à n'être que de simples raccourcis de langage, dans la mesure où elles ne possèdent précisément aucune contrepartie dans la réalité, définie par sa matérialité perçue. Une telle théorie a des implications instrumentales : tous les énoncés qui n'ont pas de contenus descriptifs ni ne sont susceptibles de recevoir une vérification expérimentale, sont considérés comme fournissant seulement une aide provisoire dont on pourra, grâce à une analyse judicieuse de leur usage linguistique, finir par se passer.

Dans le cas où l'on tente d'envisager la notion d'esprit de façon nominaliste, une théorie conséquente doit soutenir qu'il est possible d'en construire un modèle qui soit à la fois descriptible en termes finis et dont on puisse justifier la validité par expérience. C'est pourquoi le modèle computationnel se présente de lui-même comme un modèle nominaliste et c'est pourquoi aussi la justification philosophique du modèle computationnel qui fait l'objet de l'intelligence artificielle, souvent d'un raffinement extrême, découvre ou redécouvre certains thèmes rémanents de la tradition nominaliste[402].

121. Nominalisme et fonctionnalisme

La justification nominaliste au modèle computationnel consiste à souscrire à trois postulats[403]. Premièrement, les systèmes formels sont les paradigmes des systèmes représentationnels. Deuxièmement, le type de traitement que les systèmes représentationnels effectuent est de type calculatoire, au sens de la thèse de Church. Troisièmement, du fait du parallèle strict rencontré dans les systèmes formels entre syntaxe et sémantique, la référence, c'est-à-dire la possibilité d'une dénotation sémantique à partir d'un système régi exclusivement par des règles syntaxiques, doit s'expliquer par ce parallélisme. Ces trois postulats définissent, dans le cadre du modèle computationnel de l'esprit, une doctrine appelée “fonctionnalisme”.

---

[402] C'est le cas de l'hypothèse d'un “langage de la pensée”, remise en circulation par Fodor (cf. **J. Fodor**, *The Language of Thought*, Harvard University Press, Harvard, 1975) et qui est déjà présente chez Hobbes et plus loin encore, chez Ockam. Cf. **Ockam**, *Somme de Logique*, Première partie, chapitres 15-16, T. E. R., Mauvezin, 1988, pp. 51-59.

[403] C'est ce que fait remarquer D. Andler dans “Representations : beyond the pro and con”, manuscrit non-publié.

Le fonctionnalisme est une doctrine qui vise à définir le mental par le biais de termes non-mentaux ayant un statut objectif. Le paradigme des termes non-mentaux commun à toutes les versions du fonctionnalisme[404] sont les états des machines de Turing dont on postule le statut objectif, c'est-à-dire non spécifiquement "mental". Examinons les trois postulats de la démarche fonctionnaliste.

121. 1. Les systèmes formels comme paradigmes des systèmes représentationnels

On a vu [405] que concevoir les systèmes formels comme le paradigme de toute représentation provient d'un héritage mathématique qui a connu, de Frege à Hilbert, une impulsion décisive au cours de ce siècle : le formalisme hilbertien a pour but de représenter les énoncés mathématiques en les traduisant en langage logique. Dans le cas qui nous occupe, celui dans lequel la machine de Turing joue le rôle d'un modèle computationnel de la pensée, c'est l'esprit lui-même qui joue le rôle de système formel, c'est-à-dire d'une instance algébrique qui est descriptible dans un langage symbolique.

On a vu que cette définition de la représentation comme traduction paraissait trop restrictive. Si en effet la représentation est une traduction, alors on a tendance à considérer que ce que l'esprit - assimilé à un système formel de traitement de l'information - reçoit de l'extérieur est *déjà* de l'ordre du symbolique : il n'y a en effet de traduction qu'entre des symboles et non entre une réalité et un symbole. Aussi le monde extérieur ne serait perçu qu'en tant qu'il est déjà traité comme un symbole. C'est ce que fait remarquer F. Rastier quand il dit [406]:

---

[404] N. Block en définit trois grandes espèces, qu'il appelle "L'analyse fonctionnelle" qui consiste à décomposer un système en ses composants premiers, le "Fonctionnalisme computo-représentationel" qui consiste à tenter de fournir l'équivalent d'un programme informatique pour les états mentaux et le "Fonctionnalisme métaphysique" qui tente de définir la nature propre d'un état mental. Cf. **N. Block**, "What is Functionalism ?" in [**N. Block ed.**, *Readings in the Philosophy of Psychology*, I, Harvard University Press, 1980], pp. 171-172.

[405] Cf. Première partie, chapitre 1.

[406] **F. Rastier**, *Sémantique et recherches cognitives*, Presses Universitaires de France, Paris, 1991, p. 38.

«Ce paradigme met en œuvre une conception traductionniste du sens : le sens d'un symbole est sa traduction en d'autres symboles, ou plus précisément ce qu'il y a de commun avec sa traduction. On objectera alors que la représentation d'un objet par un symbole, et la représentation d'un symbole par un autre n'ont rien de commun. N'empêche cette différence est souvent négligée, comme si tous les objets du monde étaient eux-mêmes des symboles : dans l'ontologie cognitiviste spontanée, ils en partagent bien des caractéristiques, comme la discrétion et l'identité à soi».

Rien ne dit pourtant que le monde et les objets qui s'y trouvent soient spontanément descriptibles selon un format qui peut en faire des symboles : il n'y a aucune raison de considérer que le monde soit écrit sous forme de symboles [407]. C'est la notion de "transducteur" qui sert habituellement à décrire le passage réciproque[408] de l'objet au symbole. Voici comment Pylyshyn la définit [409]:

«En transformant certaines classes d'états physiques de l'environnement en états de l'appareil ayant un sens computationnel, le transducteur accomplit une conversion assez particulière : il convertit des événements physiques arbitraires du point de vue computationnel en événements computationnels».

Cette conversion s'opère par le biais d'une fonction qui doit avoir, pour Pylyshyn trois caractéristiques [410]: premièrement, le domaine de la fonction doit être descriptible en termes physiques; deuxièmement, l'ensemble d'arrivée de la fonction doit être constitué de symboles discrets, atomiques et computationnellement accessibles; troisièmement, cette transformation ne peut s'opérer que selon des principes physiques. Deux difficultés apparaissent alors.

Premièrement, la fonction des transducteurs est d'assurer le passage réciproque du formel au physique, c'est-à-dire d'attribuer un sens physique à des

---

[407] Cet aspect directement symbolique de la perception semble héritée de Leibniz. Couturat fait remarquer que pour Leibniz, «La nature est le produit d'une logique divine, de ce calcul immense qui est la création; elle est pour nous une admirable machine à calculer, car elle nous fournit, tout faits, les résultats de calculs qui dépassent la portée de notre entendement». **L. Couturat**, *La logique de Leibniz,* Félix Alcan, Paris 1901, p. 256. Leibniz justifie ainsi le caractère déjà intellectualisé de la perception; c'est ce qu'il appelle les "petites perceptions". Cf. **Leibniz**, *Nouveaux Essais*, II, 9.

[408] Ce point important est souligné par Haugeland qui distingue les transducteurs de sortie et les transducteurs d'entrée. Cf. **J. Haugeland**, "Semantic Engines : an Introduction to Mind Design" in [**J. Haugeland ed.**, *Mind Design*, op. cit.], p. 29.

[409] **Z. Pylyshyn**, *Computation and Cognition, Toward a Foundation for Cognitive Science,* A Bradford Book, MIT Press, Cambridge, Massachussetts, 1984, p. 152

[410] **Z. Pylyshyn**, *Computation and Cognition, Toward a Foundation for Cognitive Science,* op. cit., p. 154.

états formels internes et de traduire une inférence logique interne en une cause physique. Mais en tant qu'ils font partie de la machinerie matérielle des organismes vivants, les transducteurs sont *eux-mêmes* des systèmes matériels gouvernés par la causalité et non par l'implication logique[411]. Aussi est-ce le statut logique du transducteur qui fait difficulté car, pour pouvoir jouer le rôle de passage entre le registre physique et le registre logique, on ne peut que *postuler* le fait que le transducteur modélise adéquatement des informations physiques en informations logiques. Comme le fait remarquer Rosen [412]:

> **«Nous pouvons seulement faire l'hypothèse qu'*il existe* un ensemble […] de transducteurs qui vont du monde matériel au domaine des entrées sur des rubans [de machine de Turing]».**

Cette hypothèse, comme le remarque encore Rosen, est liée à l'adoption d'un présupposé réductionniste et ne va pas de soi. Elle évite en particulier de se poser la question du passage du continu physique au discret logique. La difficulté liée à la notion de transducteur est donc que le transducteur doit être capable à la fois de discrétiser le stimulus physique qu'il reçoit de l'extérieur et de produire également du continu physique comme réponse dans l'environnement. Or la façon dont s'opère cette double transformation reste obscure bien qu'elle soit absolument nécessaire pour rendre compte de l'interaction entre un individu humain et son milieu[413].

Deuxièmement, Pylyshyn reste peu précis sur le type et l'usage des principes physiques en question. A quelles lois physiques fait-il référence exactement ? Et à quel niveau de la description physique fait-il allusion ? Est-ce aux lois macroscopiques ou microscopiques ? Comme Penrose l'a remarqué, la distinction qu'il faut faire entre deux types de lois physiques, quantique et classique, a des conséquences importantes sur la définition de ce que pourrait être

---

[411] **R. Rosen**, "Processes and Natural Laws", in [*The Universal Turing machine*, **R. Herken ed**., op. cit.] p. 531 : «Les instruments d'encodage ou transducteurs sont aussi des systèmes matériels gouvernés par la causalité et non par des implications».

[412] **Rosen R.**, "Processes and Natural Laws", in [*The Universal Turing machine*, **R. Herken ed**., op. cit.] p. 532.

[413] Par exemple, la possibilité d'un échange verbal repose sur cette transformation : le locuteur discrétise une onde sonore continue quand il écoute et émet des ondes continues quand il répond.

un transducteur [414]:

> **«On sait déjà que certaines cellules nerveuses sont sensibles à un seul stimulus quantique (par exemple, sur la rétine d'une grenouille, où un simple photon peut déclencher un signal nerveux macroscopique). Pour ces sortes de cellules, l'effet consiste à convertir quelque chose qui doit apparemment être traité de façon quantique en quelque chose qui doit apparemment être traité de façon classique. Pour le moment, nous n'avons pas de théorie adéquate pour cela. La supposition que je fais est que pour certaines classes de cellules, ou peut-être pour des transmetteurs chimiques, dans le cerveau, ces sortes d'allées et venues quantique / classique sont un des composants essentiels de notre pensée».**

Ainsi, pour pouvoir interpréter les états physiques en termes de propositions formelles susceptibles de jouer le rôle d'entrées pour une machine de Turing donnée, faudrait-il déjà avoir à notre disposition une théorie physique du passage du niveau quantique au niveau classique. C'est seulement une fois cette théorie constituée que pourrait se poser la question de l'existence de règles physiques qui puissent cognitivement justifier la notion de transducteur. Mais il s'agit là d'une pure hypothèque sur l'avenir qui n'a aucune confirmation dans les faits. Pylyshyn semble donc spéculer sur un état de la physique à venir, état dont personne n'a la moindre idée, comme le fait remarquer Penrose.

121. 2. Le traitement calculatoire des données

On a vu que la thèse de Church était une thèse que l'on pouvait qualifier à bon droit de mathématique. Mais on a vu que pour lui donner un sens dans le cadre de la justification du modèle computationnel de l'esprit, il fallait faire usage de la thèse de Turing, qui elle, n'était pas seulement de nature mathématique. Aussi, tant que l'on soutient que le type de traitement que les systèmes représentationnels effectuent est de type calculatoire, au sens de la thèse de Church, *on ne s'est pas encore engagé* sur le terrain cognitif lui-même. La thèse de Church, universellement acceptée aujourd'hui par les mathématiciens et les logiciens, ne pose pas de problèmes majeurs en elle-même parce qu'elle n'est pas, en fait, de nature cognitive.

---

[414] **R. Penrose**, "On physics and mathematics of thought" dans [*The Universal Turing machine*, **R. Herken ed**., op. cit.] p. 519.

121. 3. Le parallèle entre syntaxe et sémantique

Le troisième postulat considère que la relation du système cognitif à la réalité externe peut être expliquée par le même type de moyens que ceux que l'on utilise dans les systèmes formels quand on montre en quel sens et selon quelles règles un système formel peut avoir des interprétations. Comme l'explique Haugeland [415]:

> **«Les systèmes formels (et les ordinateurs) peuvent être plus que de simples jeux parce que leurs marques peuvent avoir une interprétation qui les relient au monde extérieur. C'est le domaine de la sémantique et de la pragmatique. Quelque fois, nous disons que les marques dans un certain système formel *veulent dire* quelque chose - c'est-à-dire qu'elles sont les "signes", "symboles", ou les "expressions" qui se "tiennent à la place", "représentent" ou "disent" quelque chose. De telles relations mettent en rapport les marques et le monde extérieur (ce qu'ils désignent) en rendant possible leur usage pour des buts comme la mémorisation, la communication, le calcul etc».**

Dès lors, le problème de la représentation est celui de la façon de concevoir le rapport entre la syntaxe d'un système formel et sa sémantique. Comment s'opère le choix de l'interprétation des marques internes à un système formel ? Les propriétés syntaxiques munies d'une interprétation sont-elles suffisantes pour rendre compte d'une interaction avec la réalité externe ? C'est une remarque similaire que fait Haugeland quand il dit [416]:

> **«Interpréter un système formel automatique consiste à trouver un moyen d'analyser ses sorties (à leur assigner un sens) de sorte qu'ils fassent sens à la lumière des entrées antécédentes du système et de ses sorties. Dans le cas particulier des systèmes logiques et mathématiques, cela suffit si les sorties sont vraies au sens de la consistance; et cela peut être garanti par le fait d'avoir seulement des axiomes vrais et des règles qui préservent la vérité. Dans des systèmes plus ambitieux, cependant, qui incluent tout système ayant une aspiration à l'intelligence artificielle, la vérité n'est pas une condition suffisante pour rendre compte du sens des marques de sortie; beaucoup d'autres considérations importent également. C'est pourquoi il n'y a pas de raison de croire que la préservation de la vérité soit la seule condition, ou même la condition plus importante, qui doive être imposée aux règles du système».**

Il y a donc là comme une aporie du sens dans le cadre d'une théorie computationnelle de l'esprit qui provient de la séparation qu'elle instaure entre syntaxe et sémantique. Cette séparation radicale recoupe celle qui était apparue

---

[415] **J. Haugeland**, "Semantic Engines : an Introduction to Mind Design" dans [**J. Haugeland ed.**, *Mind Design*, op. cit.], p. 21.
[416] **J. Haugeland,** "Semantic Engines : an Introduction to Mind Design" dans [**J. Haugeland ed.**, *Mind Design*, op. cit.], p. 31.

dans l'analyse des deux premiers postulats : elle est analogue à la séparation qui existait dans le premier postulat entre registre logique et registre physique dans la définition du transducteur; de même, dans le cas du deuxième postulat, la thèse de Church devait être "augmentée" de la thèse de Turing pour pouvoir jouer un rôle de principe cognitif.

Le fonctionnalisme se heurte donc de front à la difficulté de fonder un rapport entre le niveau logique capable de modéliser le "mental" et un niveau physique, d'une autre nature que le niveau logique. Cette difficulté provient de l'hypothèse de départ, qui pose en principe la différence entre le corps et l'esprit. C'est la raison pour laquelle le rapport du fonctionnalisme au physicalisme est de nature aporétique.

Tout se passe comme si la notion de représentation jouait en fait sur trois plans différents - le plan intuitif du sens, le plan physique et le plan formel - mais que ces trois plans n'entretenait pas le même type de rapport les uns avec les autres. Il faudrait distinguer le plan du sens et le plan physique d'une part et le plan formel qui s'opposerait aux deux premiers. Nous reviendrons sur les rapports qu'entretiennent ces trois plans.

122. Fonctionnalisme et physicalisme

Comme nous l'avons déjà remarqué en suivant N. Block, le fonctionnalisme a la particularité, quand il est utilisé comme argument philosophique[417], de servir à la fois à justifier le physicalisme et à l'infirmer. En fait, comme nous l'avons déjà souligné, la difficulté provient de ce que le formalisme de la machine de Turing appliqué au cas de l'esprit n'opère pas une contrainte assez forte pour assurer une relation univoque entre le niveau logique et le niveau physique. On se trouve alors devant le problème suivant : soit une même machine de Turing peut se réaliser dans un nombre indéfini d'entités physiques, soit une même entité physique peut recevoir un nombre indéfini de descriptions

---

[417] Nous laissons de côté l'analyse du fait que, dans la littérature, le fonctionnalisme se présente à la fois comme programme de recherche et comme thèse philosophique. Cette si curieuse dualité ne semble pas avoir retenu l'attention des chercheurs. On pourrait cependant légitimement se demander comment un programme de recherche, s'il est bien de nature scientifique, pourrait servir à justifier une position philosophique quelle qu'elle soit.

sous forme de machines de Turing.

Dans le cas de la modélisation des états mentaux par le biais du formalisme de la machine de Turing, on se trouve donc pris entre deux options qu'il faut également rejeter. La première consiste à accorder des états mentaux à toute entité susceptible d'être modélisée par une fonction récursive. Mais on voit que cette version du fonctionnalisme est beaucoup trop "libérale" : si l'on reprend l'exemple de N. Block, il est possible de modéliser l'activité économique d'un pays en termes de fonctions récursives et pourtant personne ne voudrait accorder un esprit à ce type d'entités[418]. La seconde consiste à n'accorder des états proprement mentaux qu'à l'espèce humaine, en excluant tout autre entité physique. Mais on voit mal alors pour quelles raisons l'attribution du qualificatif "mental", s'il est associé à un modèle en termes de machines de Turing, serait limité au substrat physique des êtres humains puisque le même formalisme peut modéliser tout autre chose.

Aussi comme le fait remarquer N . Block [419]:

> **«[…] toute tentative visant à formuler une description fonctionnelle possédant des caractérisations physiques sous forme d'entrées et de sorties conduira inévitablement à exclure certains systèmes ayant un esprit ou à inclure des systèmes sans esprit».**

On voit donc que le problème du rapport entre le fonctionnalisme et le physicalisme se pose parce que le problème de la distinction entre un niveau mental et un niveau physique n'a pas été résolu mais que celle-ci a été présupposée.

De ce point de vue, les nominalistes seraient, dans le contexte du modèle computationnel de l'esprit, des dualistes "honteux", tandis que l'attitude platonicienne y verrait au contraire l'occasion de corroborer son attitude dualiste. Il faut revenir à la façon dont les états mentaux sont identifiés à des états de machine de Turing pour essayer de comprendre où exactement se situe le débat.

---

[418] Cf. **N. Block**, "Troubles with Functionalism" in [**N. Block ed.**, *Readings in the Philosophy of Psychology*, I, Harvard University Press, 1980], p. 294.

[419] Cf. **N. Block**, "Troubles with Functionalism" in [**N. Block ed.**, *Readings in the Philosophy of Psychology*, I, Harvard University Press, 1980], p. 295.

2. L'opposition sur la nature de l'abstraction

Le débat sur la façon dont il est possible d'appréhender les états mentaux par le biais de machines est révélateur de l'opposition entre le point de vue platonicien et le point de vue nominaliste.

21. Critique du nominalisme par Gödel

Une conception nominaliste de la description de l'intelligence paraît à Gödel avoir des fondements chez Turing à qui il attribue une "erreur philosophique"[420]:

> **«*Une erreur philosophique dans l'œuvre de Turing*. Turing dans son *1937*, page 250 [dans** Davis M. **p. 136]** [*]**, donne un argument qui est supposé montrer que les procédures mentales ne peuvent pas aller au-delà des procédures mécaniques. Quoiqu'il en soit, cet argument n'est pas concluant, parce qu'il dépend de la supposition selon laquelle un esprit fini est seulement capable d'un nombre fini d'états distincts. Ce que Turing laisse entièrement de côté, c'est que *l'esprit, dans son fonctionnement, n'est pas statique mais se développe constamment*, c'est-à-dire que nous comprenons de mieux en mieux les termes abstraits au fur et à mesure que nous les utilisons. Il peut exister des méthodes systématiques permettant d'actualiser un développement, méthodes qui peuvent former une partie de la procédure. Donc, bien qu'à chaque étape le nombre et la précision des termes abstraits à notre disposition soit fini, ces deux caractéristiques (et donc aussi le nombre d'états distincts de l'esprit comme le veut Turing) peuvent converger vers l'infinité au cours de l'application de la procédure. Il est à noter que quelque chose de ce type semble se produire dans le processus qui consiste à former des axiomes de l'infini de plus en plus puissants en théorie des ensembles. Ce processus, cependant, est loin d'être suffisamment compris pour former une procédure bien définie. On doit admettre que la construction d'une procédure bien définie qui pourrait être véritablement menée à bien (et produirait des fonctions numériques non-récursives) demanderait d'accomplir**

---

[420] **K. Gödel**, "Some remarks on the undecidability results" (1972a), republié dans [**K. Gödel**, *Collected Works*, t. II, op. cit.], p. 306.

[*] C'est Gödel qui renvoie à l'anthologie éditée par Davis dans lequel se trouve l'article de Turing. Gödel fait sans doute allusion au passage suivant : «Le comportement du calculateur [*computer*] à n'importe quel moment est déterminé par les symboles qu'il [*he*] observe et son [*his*] "état d'esprit" à ce moment. Nous pouvons supposer qu'il existe une borne *B* au nombre de symboles ou de cases que le calculateur peut observer à chaque moment du temps. S'il veut en observer plus, il doit exécuter plusieurs observations successives. Nous ferons aussi l'hypothèse que le nombre d'états d'esprit que l'on a besoin de prendre en considération est fini. Les raisons de cet état de fait sont les mêmes que celles qui restreignent le nombre des symboles. Si nous admettions une infinité d'états d'esprit, certains d'entre eux seraient "arbitrairement proches" et seraient confondus. Encore une fois, cette restriction n'affecte pas gravement le calcul puisque l'usage d'états d'esprit plus compliqués peut être évité en écrivant plus de symboles sur le ruban». **A. M. Turing**, (1939), ""On Computable Numbers …"", Proceedings of the London Mathematical Society, 42 : 230-265; republié dans [**M. Davis ed.**, *The Undecidable* Raven Press, Hewlett, New-York], p. 136. Nous reviendrons sur ce texte au paragraphe 22.

un progrès substantiel dans notre compréhension des concepts fondamentaux des mathématiques».

Pour Gödel, la mémoire peut s'infinitiser parce qu'elle est *omnisciente*. Cette omniscience tient à la façon de concevoir l'opération non-mécanique consistant à isoler des éléments de la mémoire. L'opération en question est de nature *sémantique* : la compréhension par l'esprit de situations de plus en plus complexes exige l'utilisation de termes abstraits qui apparaissent comme des résumés représentant de multiples états d'esprit individuels[421]. L'idée de convergence infinie des étapes successives de la théorie axiomatique des ensembles qui sert d'exemple à Gödel fait donc appel à la notion d'une dynamique interne de l'esprit produisant une *abstraction*, dont les décisions, successives sans être susceptibles d'être entièrement assignables à une règle récursive dans leur totalité, permettent de forger de nouveaux axiomes plus puissants. Gödel justifie cette capacité de "représentation abstraite" non-mécanique par le platonisme : la plus grande complexité des axiomes inventés implique d'entrer dans des états d'esprit eux-mêmes plus complexes. *Il y a donc une efficacité sémantique de l'objet mathématique sur les états mentaux de l'être humain qui l'appréhende*. La représentation abstraite est le résultat mental de cette "efficacité de l'objet". A ce sujet, R. Gandy fait remarquer que la capacité de représentation abstraite telle que Gödel la conçoit peut être décrite par le biais de la notion de *Gestalt* [422]:

«L'intelligence non-mécanique verrait, pour ainsi dire, l'état *x* comme une *Gestalt* et effectuerait, par le biais d'une abstraction, des déterminations globales qui ne

---

[421] L'inspiration de Gödel est sans doute, ici encore, autant leibnizienne que platonicienne : «On peut même dire que les sciences s'abrègent en s'augmentant, qui est un paradoxe très véritable, car plus on découvre des vérités et plus on est en état d'y remarquer une suite réglée et de se faire des propositions toujours plus universelles dont les autres ne sont que des exemples ou corollaires, de sorte qu'il se pourra faire qu'un grand volume de ceux qui nous ont précédé se réduira avec le temps à deux ou trois thèses générales. Aussi plus une science est perfectionnée, et moins a-t-elle besoin de gros volumes, car selon que ses Éléments sont suffisamment établis, on y peut trouver tout par le secours de la science générale ou de l'art d'inventer». **Leibniz**, "Discours touchant la méthode de la certitude et l'art d'inventer", *die Philosophischen Schriften*, ed. Gerhardt, Olms Verlag, Hildesheim, 1890, Band VII, p. 180.

[422] **R. O. Gandy**, "Church's thesis and principles for mechanisms" in [*The Kleene Symposium*, **J. Barwise**, **J. J Keisler** et **K. Kuchen eds.**, Amsterdam, North-Holland Publ. Co. (1980) : 123-145], p. 146.

pourraient pas être obtenues par des méthodes locales».

Il y aurait, ainsi, dans la notion de *Gestalt*, un surcroît sémantique non calculable, que Gödel rapporte, en platonicien, à la valeur intrinsèque de l'objet mathématique. A ma connaissance, ni Gödel, ni Gandy après lui, n'ont essayé d'aller plus loin dans la description rigoureuse de cette représentation abstraite par *Gestalten* produites par l'objet mathématique. On peut supposer qu'elles impliquent des contraintes de connexité qui constituent des éléments épars en des touts cohérents[423]. Mais il suffit pour nous de remarquer qu'elle est justifiée par une conception platonicienne de l'abstraction puisque c'est sur ce point que Gödel et Turing divergent.

Contrairement à ce que soutient Gödel, Turing s'est posé à lui-même l'objection qu'il lui adresse touchant la possibilité d'une prise en compte par l'esprit d'un nombre infini de ses états. Il a finalement rejeté cette objection grâce à un certain nombre d'arguments de nature nominaliste que nous allons étudier maintenant.

22. Les arguments nominalistes de Turing

Turing, dans "On Computable Numbers …", ne remet pas en question la possibilité de l'existence d'un nombre infini d'états mentaux et s'interroge seulement sur la façon dont il serait possible de les appréhender. Turing fait remarquer de ce point de vue que la mémoire humaine lui paraît être «nécessairement limitée»[424] et c'est ce qui pour lui est un *fait* qui justifie la définition des réels calculables en termes de machines. J. Webb a précisé en ces termes l'argumentation de Turing [425]:

«Le cœur de son argument était une nouvelle analyse abstraite de ce que l'on entend logiquement par "se souvenir effectivement" des choses ayant rapport au calcul, tels que

---

[423] C'est ce que l'école de la *Gestalt Theorie* appelle «l'ensembléité» des éléments formant un tout. Cf. **W. Köhler**, *Psychologie de la forme*, Gallimard, Paris, 1964, p. 175.

[424] **A. M. Turing**, 'On Computable Numbers with an Application to the Entscheidungsproblem', op. cit., republié dans [**M. Davis ed.**, *The Undecidable*, op. cit.], , p. 117.

[425] **J. C. Webb**, "Introductory note to 1972a" in [**K. Gödel**, *Collected Works*, vol. II, Oxford University Press, Oxford, 1990], p. 302.

les symboles ou combien de fois on doit exécuter une sous-routine : pour ce faire, on doit être capable de passer d'un état discernable à un autre, *que vous soyez un humain ou une machine*. Nous faisons évidemment l'hypothèse que les "états d'esprit" peuvent aussi avoir des souvenirs qui dépassent les rêves les plus fous des machines, mais les seuls qui soient pertinents pour le calcul effectif sont ceux dans lesquels vous êtes placés par les symboles et les processus qui apparaissent au cours du calcul».

Si l'on suit l'argument de Webb, ce sont les symboles qui nous placent dans des états d'esprit particuliers et non les actes de l'esprit qui se matérialisent sous forme de symboles. Le fond de l'analyse de l'acte de calcul tel qu'il est présenté par Turing est donc qu'il y a une *efficacité intrinsèque aux symboles* qui limite le champ de la pertinence des symboles choisis sans limiter *a priori* le nombre des symboles pris en considération : seuls les symboles discernables doivent être pris en considération, quel que soit leur nombre. C'est cette efficacité prêtée aux symboles qui constitue l'acte nominaliste dans l'analyse du calcul par Turing. Gödel ne prend pas en compte cette efficacité prêtée au langage puisqu'il la rapporte directement à l'objet. C'est précisément sur l'appréhension d'un nombre infini de symboles que la différence apparaît clairement. Turing fait remarquer [426]:

«Si nous admettions un nombre infini d'états d'esprit, certains seront arbitrairement proches et seront confondus. Encore une fois, la restriction n'affecte pas sérieusement le calcul, puisque l'utilisation d'états d'esprit plus compliqués peut être remplacé par l'écriture de plus de symboles sur le ruban».

Cette façon de poser le problème tend à montrer que la difficulté ne vient pas du nombre infini des états mais plutôt de leur complexité. Or on peut pallier la difficulté liée à la *complexité* des états d'esprit en *augmentant* le nombre des états de la machine, quel que soit ce nombre. L'exemple le plus probant de ce point de vue est celui des machines universelles. En effet, n'importe quelle machine, même celles dont les états internes sont très complexes, peut être imitée par une machine universelle, parce que le problème de la complexité des états peut toujours être contourné par le biais d'un surcroît d'écriture. La machine universelle apparaît donc comme le paradigme de cette efficacité symbolique qui compense toujours la

---

[426] **A. M. Turing**, 'On Computable Numbers with an Application to the Entscheidungsproblem', op. cit., republié dans [**M. Davis ed.**, *The Undecidable*, op. cit.], , p. 136.

complexité croissante des machines à simuler malgré son nombre fini d'états[427].

Dès lors, le raisonnement de Turing est qu'il n'y a pas besoin de supposer une efficacité de l'objet sur l'esprit parce que l'écriture en dispense. Turing le dit d'ailleurs lui-même [428]:

> **«Il est toujours possible pour le calculateur de faire une pause dans son travail, de partir et de tout oublier à son sujet, puis de revenir plus tard et de s'y remettre».**

L'être humain peut donc oublier ce qu'il faisait et changer entièrement les états de son esprit, il n'en reste pas moins capable, grâce aux instructions qu'il a écrites et conservées hors de lui, de reprendre ensuite son calcul pour retrouver immédiatement les états d'esprit finis qui étaient les siens.

La différence entre Turing et Gödel tient donc à la façon dont ils interprètent le processus de l'abstraction. La perspective platonicienne renvoie à ce qui est sémantiquement *compréhensible* : l'objet vient en premier par rapport à la pensée qui en tire pour elle-même la possibilité d'une complexification progressive de ses états internes. La perspective nominaliste renvoie à ce qu'il est possible de *traduire* par l'écriture de états internes du calculateur : la possibilité pour la pensée de s'extérioriser par l'écriture confère à cette dernière un rôle moteur dans la possibilité d'une réflexion de l'esprit sur lui-même. C'est ce renversement de perspective qui fait passer, dans la perspective nominaliste, l'objet en second par rapport à la pensée et qui constitue le fond de la différence entre les deux points de vue. On peut dire alors que ce que le point de vue nominaliste reproche au point de vue platonicien, c'est de croire illusoirement à une intuition immédiate - sans langage - de l'objet, tandis que le point de vue platonicien reproche au point de vue nominaliste de faire plus porter son attention

---

[427] Turing dit dans sa conférence de 1947 : «Ainsi la complexité de la machine à imiter est concentrée dans le ruban et n'apparaît en aucune façon dans la machine universelle elle-même». **A. M. Turing**, 'Lecture to the London Mathematical Society', 20 february 1947; republié dans [**B. E. Carpenter** et **R. W. Doran eds**., *A. M. Turing's ACE report of 1946 and other papers*, MIT Press, Cambridge, Massachussetts, 1986], p. 112.

[428] **A. M. Turing**, 'On Computable Numbers with an Application to the Entscheidungsproblem', Proceedings of the London Mathematical Society, 42 : 230-265; republié dans [**M. Davis ed.**, *The Undecidable*, op. cit.], p. 139.

sur le fonctionnement de l'esprit que sur la source objective de ses connaissances, c'est-à-dire de sombrer dans le *psychologisme*.

Dans le cas présent, cette source objective est le continu des réels. La maîtrise de son calcul passe par une analyse de la représentation que l'esprit s'en fait, analyse qui a produit le modèle computationnel de l'esprit et ses interprétations dualistes du processus d'abstraction. La réponse au problème du calcul du continu des réels par le biais du modèle computationnel nécessite-t-elle une conception dualiste de l'abstraction ?

Que celle-ci soit justifiée par un raisonnement platonicien ou par un raisonnement nominaliste, les deux conceptions partent d'un présupposé commun qui consiste à envisager l'esprit soit comme une entité radicalement différente de la réalité matérielle du cerveau soit comme n'entretenant avec cette réalité matérielle qu'un rapport déjà abstrait d'ordre linguistique. On peut se demander alors s'il peut y avoir une critique possible du platonisme à l'égard du nominalisme, et *vice versa*, puisque leur opposition semble être de nature dialectique et se fonder, en fait, sur le même présupposé. Le conflit ne semble pouvoir se dissoudre que si l'on en revient à son fondement, la distinction de l'esprit et du corps, ou ses substituts plus abstraits, distinction acceptée par les deux théories philosophiques en présence[429]. De ce point de vue, on peut se demander si le dualisme de ces interprétations ne provient pas du fait qu'elles ne reconnaissent pas la nécessité d'une médiation entre l'aspect proprement logique du modèle et son aspect physique. Or l'interprétation psychologique que nous avons donnée du jeu de l'imitation a précisément tenté de montrer qu'un modèle computationnel de l'esprit n'était devenu possible qu'à partir du moment où une médiation de ce type s'était constituée. Telle qu'elle se manifestait dans "Computing Machinery and Intelligence", cette médiation reposait sur un usage d'images qui permettait de constituer le jeu de l'imitation en image du corps.

---

[429] Le même Kurt Gödel avait eu cependant, à la mort de Turing, une réaction qui aurait pu laisser supposer une attitude moins franchement dualiste dans les rapports de l'esprit et du corps. Hao Wang rapporte à ce propos le fait suivant : « *1956*. En août, Kreisel et moi sommes allés prendre le thé chez les Gödel (qui habitaient Linden Lane). Quand on mentionna le suicide de Turing, Gödel demanda s'il était marié. Quand on lui eut fourni une réponse négative, il dit : "Peut-être voulait-il se marier, mais ne le pouvait pas"». **W. Hao**, *Reflections on Kurt Gödel*, op. cit., p. 120.

Ainsi, entre le logique et le physique, notre analyse du jeu a tenté de montrer que la constitution du modèle computationnel nécessitait de faire intervenir une *fonction dissimulatrice* grâce à laquelle il devenait possible d'articuler les aspects physiques et psychiques propres au domaine du mental.

Il faut donc tenter de reprendre à nouveaux frais la distinction entre le corps et l'esprit et voir s'il ne serait pas possible d'interpréter leur rapport par le biais d'une théorie non-dualiste de la représentation qui envisagerait autrement le problème de l'abstraction.

Chapitre II

Le rôle du langage dans l’interprétation nominaliste du modèle computationnel de l’esprit

Dans l’acte de calcul d’un nombre réel, l’esprit en tant qu’il reconnaît l’existence du continu formé par les réels ne se réfléchit pas comme composé d’états discrets mais comme identique à son objet, c’est-à-dire comme continu[430]. L’esprit est ainsi capable d’appréhender l’ensemble des suites rationnelles et irrationnelles et d’envisager cet ensemble comme formant un tout. Dans l’interprétation nominaliste de l’acte de calcul telle qu’elle a été mise au jour par J. Webb, c’est parce que le langage logique utilisé par la machine opère une contrainte sur l’esprit que ce langage décompose l’esprit en états discrets. Dès lors, le langage en question apparaît comme d’une autre nature que l’acte de reconnaissance de l’existence du continu par l’esprit : il a un statut *d’abstraction* par rapport à un acte de reconnaissance qui est immédiat et lui est antérieur. Se conformer à l’ordre abstrait propre au langage logique implique donc d’introduire l’idée d’une succession d’états discrets pour appréhender le continu des nombres réels par le biais de leur développement décimal.

Si l’on suit l’interprétation nominaliste donnée par J. Webb de l’argumentation de Turing touchant le rapport entre les états d’une machine et les états de l’esprit, les seuls états d’esprit pertinents pour l’acte de calcul des suites décimales des nombres réels par machine de Turing sont ceux dans lesquels le calculateur est placé *par* les symboles utilisés : ce sont les symboles abstraits qui nous placent dans des états

[430] Cette question a été abordée dans la première partie, chapitre 3, § 122. 22.

d'esprit spécifiques et non l'inverse. L'usage de ces symboles aurait ainsi une influence sur ce qu'il faut concevoir comme pertinent du point de vue des états de l'esprit, c'est-à-dire opérerait une contrainte d'ordre sémantique sur l'esprit. Comment concevoir cette contrainte sémantique que semble opérer l'usage d'un langage logique ?

Il y a là une difficulté réelle parce qu'il semble à première vue difficile de concevoir que le simple usage d'un langage, fût-il logique, puisse opérer des contraintes sur la sémantique du langage en question. Il est donc nécessaire, pour comprendre le sens de cet argument, de commencer par un bref rappel de quelques notions linguistiques. On se reportera ensuite au cadre du jeu de l'imitation parce que c'est dans ce cadre que sont explicitées les contraintes produites par l'usage d'un langage logique.

1. Quelques rappels linguistiques

Tel qu'il se donne, le jeu de l'imitation instaure une situation de *discours* au sens que Benveniste a donné à ce terme en l'opposant à la langue. Le jeu consiste en effet en un échange de questions et de réponses prenant la forme d'un dialogue. C'est ce dialogue qui se trouve être soumis à des règles particulières. Il faut, avant d'en venir au jeu de l'imitation proprement dit, analyser la différence entre le discours et la langue et les conséquences qu'elle implique pour notre propos. Ces rappels ont essentiellement pour but de justifier la position première de la langue dans la hiérarchie des systèmes sémiotiques ainsi que de montrer les différences existant entre deux systèmes sémiotiques particuliers, celui de la parole et celui de l'écriture.

11. Différence entre discours et langue

Partons d'un exemple emprunté à E. Cassirer [431]. Cassirer fait remarquer que lorsque l'on observe l'étymologie en grec et en latin d'un nom comme la lune, on voit que le terme grec [*mên*] désigne "ce qui mesure" tandis que le synonyme latin [*luc-na*] désigne "ce qui brille"[432]. Ce grâce à quoi un objet est nommé, c'est-à-dire la

---

[431] **E. Cassirer**, *La philosophie des formes symboliques*, t. I, p. 255.

[432] **E. Cassirer**, *La philosophie des formes symboliques*, t. I, p. 255 commentant cet exemple : «La construction des concepts dans le langage […] ne reçoit pas ses impulsions essentielles du

représentation qui préside à la constitution d'un mot, est le fait d'une *phrase* et non d'un signe : c'est donc par l'intermédiaire du *discours* qu'un mot de la langue se constitue[433]. Or la constitution du sens dans l'univers du discours passe par deux tropes fondamentaux : la métaphore et la métonymie. Dans notre exemple, la constitution d'un mot appartenant à la langue comme celui de lune passe par une élaboration discursive de nature métaphorique (dans le cas grec) ou métonymique (dans le cas latin). C'est ce phénomène du passage du discursif au linguistique qui doit être précisé puisqu'il permet de délimiter le domaine propre de la représentation dans le cadre de l'interprétation nominaliste du modèle computationnel. On peut donc dès à présent en décrire les deux modes fondamentaux : permettant le remplacement par une entité matérielle, la représentation d'un objet par un nom implique néanmoins, de par l'usage de symboles, une *interprétation.*

En effet, la référence à l'objet “lune” se fait par la mise en exergue d'une caractéristique attribuée à l'objet dans l'univers du discours parmi un nombre indéfini de caractéristiques possibles. Chaque langue, ou plutôt chaque groupe de langues, semble posséder ainsi une façon propre d'opérer un “renvoi” à l'objet grâce à une caractéristique discursive particulière qui permet de le dénommer. En comparant différentes langues entre elles, on parvient donc à appréhender le caractère spécifique de la construction de chaque langue (ou groupe de langues). Aussi le phénomène du “renvoi à l'objet” est-il lui-même universel : se manifestant différemment dans chaque langue ou groupe de langues, il manifeste cependant un pouvoir spécifique de l'esprit humain touchant la façon de concevoir des symboles. C'est dans le passage du discours à la langue que ce pouvoir se manifeste.

12. Différence entre sémantique et sémiotique

Le discours a pour domaine le sémantique tandis que la langue a pour domaine

---

seul monde de l'être, mais toujours aussi du monde de l'agir. Les concepts du langage se situent toujours à la limite entre l'action et la réflexion, entre l'agir et le contempler».

433 On comprend pourquoi Benveniste peut soutenir : «C'est dans le discours, actualisé en phrases, que la langue se forme et se configure.» Cf. **E. Benveniste**, *Problèmes de linguistique générale*, “Les niveaux de l'analyse linguistique”, Tome 1, p. 131.

le sémiotique [434]:

> **«Le sémiotique se caractérise comme une propriété de la langue, le sémantique résulte d'une activité du locuteur qui met en action la langue. Le signe sémiotique existe en soi, fonde la réalité de la langue, mais il ne comporte pas d'applications particulières; la phrase, expression du sémantique, n'est *que* particulière. Avec le signe, on atteint la réalité intrinsèque de la langue, avec la phrase on est relié aux choses à l'extérieur de la langue; […] Une première constatation est que le "sens" (dans l'acception sémantique qui vient d'être caractérisée) s'accomplit dans et par une forme spécifique, celle du syntagme, à la différence du sémiotique qui se définit par une relation de paradigme. D'un côté la substitution, de l'autre la connexion, telles sont les deux opérations typiques et complémentaires».**

Le paradigme est un axe permettant la substitution des formes, tandis que le syntagme est un axe qui permet la constitution d'entités formant un tout. Commençons par étudier l'axe paradigmatique qui constitue le plan sémiotique.

121. La structure de la langue

C'est le caractère *discret* des signes qui exige l'emploi d'une double relation. Comme le fait remarquer Benveniste [435]:

> **«Du fait que le entités linguistiques sont discrètes, elles admettent deux espèces de relation : entre éléments de même niveau ou entre éléments de niveaux différents. Ces relations doivent être bien distinguées. Entre les éléments de même niveau, les relations sont *distributionnelles*; entre éléments de niveau différent, elles sont *intégratives*».**

La relation d'intégration réalise un emboîtement qui va du trait phonétique à la phrase. Le trait phonétique est l'unité indécomposable présent dans tout phonème. Dans [*d*ʻ] par exemple, on reconnaît quatre traits distinctifs : occlusion, dentalité, sonorité, aspiration. Ces traits distinctifs ne sont plus décomposables mais s'intègrent dans des unités plus hautes, celles des phonèmes. Les phonèmes, sont des éléments matériels *discrets* en nombre limité et fixe pour chaque langue[436]. Les phonèmes sont à la fois constitués de traits distinctifs et intégrés dans des unités plus hautes, les signes. Par exemple, *bête* possède quatre phonèmes. A ce niveau, on peut opérer des

---

[434] **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "La forme et le sens dans le langage", p. 225.

[435] **E. Benveniste**, *Problèmes de linguistique générale*, "Les niveaux de l'analyse linguistique", Tome 1, p. 124.

[436] Pour reprendre l'exemple de Martinet : *tête* se distingue, du point de vue de la seconde articulation, de *bête*, de *fête*, etc. La seconde articulation consacre l'indépendance des composants du signifiant de la valeur du signifié : chaque composant d'un signifiant se retrouve dans d'autres signifiants.

substitutions et remplacer le phonème [*b*] par le phonème [*f*] : on obtient alors le signe *fête*. Le niveau d'intégration où un nouveau type d'unité apparaît est délimité par le sens[437]. Par exemple, /s/ dans /-al/; au niveau supérieur, /sal/ fonctionne comme intégrant de *-à manger*; *-de bains*. Pour Benveniste, le modèle de la relation intégrante est celle de la fonction propositionnelle de Russell[438]. Mais la relation intégrante définie comme fonction propositionnelle ne permet pas de décrire la nature de cette intégration, qui est, dans les langues naturelles, soit métonymique, soit métaphorique.

La relation de constitution va de la phrase au trait distinctif. La phrase ne s'intègre dans rien et ne possède que des constituants, signes, phonèmes et traits distinctifs. Aussi, la phrase est-elle composée de signes sans être elle-même un signe. C'est pourquoi une proposition ne s'intègre pas dans une autre mais précède ou suit une autre proposition. Elle n'admet qu'un type de relation, la relation *prédicative* qui n'est pas de nature intégrative. Avec la phrase, on quitte le domaine de la langue comme système de signes pour entrer dans celui de la langue comme système de communication : bref, on passe de la langue au discours.

122. La structure du discours

Toute idée doit, psychologiquement, s'investir par le discours dans une forme syntagmatique[439]. On voit donc en quoi le discours se distingue de la langue : le mot est l'unité du discours, le signe est l'unité de la langue, le premier exige d'être *compris* tandis que le second demande à être *reconnu* [440]. Le message du discours ne se réduit

---

[437] **E. Benveniste**, *Problèmes de linguistique générale*, "Les niveaux de l'analyse linguistique", Tome 1, p. 126 : «Que faut-il pour que dans ces constituants formels nous reconnaissions, s'il y a lieu, des unités d'un niveau défini ? Il faut pratiquer l'opération en sens inverse et voir si ces constituants ont fonction intégrante au niveau supérieur. Tout est là : la dissociation nous livre la constitution formelle; l'intégration nous livre des unités signifiantes».

[438] **E. Benveniste**, *Problèmes de linguistique générale*, "Les niveaux de l'analyse linguistique", Tome 1, p. 125.

[439] **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "La forme et le sens dans le langage" p. 226 : «Que l'idée ne trouve forme que dans un agencement syntagmatique, c'est là une condition première, inhérente au langage. Le linguiste se trouve ici devant un problème qui lui échappe; il peut seulement conjecturer que cette condition toujours nécessaire reflète une nécessité de notre organisation cérébrale».

[440] **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 64-65 : «Le sémiotique (le signe) doit être reconnu, le sémantique (le discours) doit être compris. la différence entre reconnaître et comprendre renvoie à deux facultés distinctes de l'esprit : celle de percevoir l'identité entre l'antérieur et l'actuel, d'une part, et celle de percevoir la signification

dès lors pas à des unités susceptibles d'être identifiées séparément [441]:

> **«[…] ce n'est pas une addition de signes qui produit le sens, c'est au contraire le sens (l'"intenté"), conçu globalement, qui se réalise et se divise en signes particuliers, qui sont les mots. En deuxième lieu, le sémantique prend nécessairement en charge l'ensemble des référents, tandis que le sémiotique est par principe retranché et indépendant de toute référence. L'ordre sémantique s'identifie au monde de l'énonciation et à l'univers du discours».**

La différence entre le sémantique et le sémiotique ne recoupe pas la tripartition logique de la syntaxe, de la sémantique et de la pragmatique. C'est la place du syntaxique qui fait problème [442]:

> **«Ces trois notions [syntaxe-sémantique-pragmatique] constituent un ensemble qui est tout autrement articulé que ce que la langue en elle-même permet de concevoir. Ensemble ou séparément, elles appartiennent exclusivement au domaine qui est, dans ma terminologie, celui du sémantique. En effet, ce qui pour le logicien est syntaxique, c'est-à-dire la liaison entre les éléments de l'énoncé, relève d'une considération qui est pour moi ambiguë, en ce sens que d'une part, ce qui est syntagmatique pour le linguiste coïncide avec ce que l'on appelle syntaxique en logique, et qui, par conséquent, se situe à l'intérieur de l'ordre du sémantique; mais d'autre part, aux yeux du linguiste, cette liaison peut être gouvernée par une nécessité purement grammaticale, qui dépend entièrement de la structure de l'idiome, qui n'est pas quelque chose d'universel, qui prend des formes particulières suivant le type de langue considérée».**

La tripartition logique provient du souci de bien distinguer la langue, les choses et les locuteurs mais elle tend du même coup à occulter l'opposition entre d'une part l'ordre *universel* de la langue et l'ordre toujours *particulier* du discours puisque ce qui est rangé sous le label du syntaxique appartient aux deux ordres à la fois. C'est précisément le problème du rapport entre ces deux façons de concevoir le langage qui fait difficulté, puisque l'une est nominaliste et l'autre ne l'est pas.

Dans la perspective de la différence entre le sémantique et le sémiotique, notons que Benveniste caractérise comme continu [443] (mais peut-être vaudrait-il mieux parler de connexité ici) le domaine du sémantique qui rend possible la référence à ce qui est

---

d'une énonciation nouvelle, de l'autre. Dans les formes pathologiques du langage, les deux facultés sont fréquemment dissociées».

441 **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 64.

442 **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "La forme et le sens dans le langage", p. 233.

443 **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 48 : «La langue se présente sous tous ses aspects comme une dualité : institution sociale, elle est mise en œuvre par l'individu; discours continu, elle se compose d'unités fixes».

en dehors du langage, à savoir le monde en tant qu'il est réel ou idéal. En revanche, la langue est composée d'unités *discrètes* qui n'ont de valeur qu'en tant qu'elles s'opposent mutuellement de façon combinatoire. Aussi les deux ordres du sémantique et du sémiotique, qui se manifestent tous les deux dans la langue, semblent-ils aussi s'exclure mutuellement [444]:

> **«Du signe à la phrase, il n'y a pas transition, ni par syntagmation, ni autrement.»**

La langue possède donc une double signifiance, par les signes d'une part et par l'énonciation d'autre part quand elle est utilisée dans un discours. C'est ce qui permet à la langue de posséder un pouvoir métalinguistique grâce auquel il est possible de désigner par la langue la langue elle-même : il est toujours possible de faire référence à l'énonciation, dans un deuxième niveau d'énonciation qui interprète le premier[445].

Du point de vue de l'énonciation, y-a-t-il une différence entre la langue telle qu'elle est parlée et telle qu'elle écrite ? Cette différence est importante pour ce qui nous concerne, puisque, dans le jeu de l'imitation, c'est bien de signes *écrits* qu'il s'agit et non directement de signes prononcés.

13. Différence entre parole et écriture

Il existe à l'évidence une différence entre la langue parlée et la langue écrite. Mais ces deux incarnations de la langue ont-elles la même fonction ? Assurément non, puisque, selon le principe de "non-redondance" tel qu'il est exprimé par Benveniste [446]:

> **«[…] deux systèmes sémiotiques de type différent ne peuvent être mutuellement convertibles».**

Comment caractériser les différences entre les deux systèmes ?

Benveniste analyse les différences entre les systèmes sémiotiques au moyen de

---

444 **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 65.

445 Benveniste y voit la raison pour laquelle la langue a la primauté sur les autres systèmes sémiotiques : la langue peut, par l'énonciation, se désigner elle-même et tous les autres systèmes sémiotiques. Cf. **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 65.

446 **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 53.

trois relations [447]: premièrement, la relation *d'engendrement* (par exemple entre l'écriture ordinaire et l'écriture sténographique ou entre la langue naturelle et la formalisation logico-mathématique); deuxièmement, la relation *d'homologie* établissant des corrélations entre deux systèmes sémiotiques (Benveniste donne pour exemple les "correspondances" de Baudelaire entre parfums, couleurs et sons et le rapport entre le geste rituel et l'écriture en Chine); troisièmement, la relation *d'interprétance* qui marque le primat de la langue sur les autres systèmes sémiotiques dans la mesure où tous les systèmes sémiotiques peuvent être décrits par la langue et que seule la langue est susceptible de se décrire elle-même. Ces trois relations permettent à Benveniste d'opérer un classement hiérarchique entre systèmes sémiotiques dans lequel la langue occupe la première place [448]:

> **«[…] la langue occupe une situation particulière dans l'univers des systèmes de signes. Si l'on convient de désigner par *S* l'ensemble de ces systèmes et par *L* la langue, la conversion se fait toujours dans le sens *S* -> *L*, jamais l'inverse. Nous avons là un principe général de hiérarchie, propre à être introduit dans la classification des systèmes sémiotiques et qui servira à construire une théorie sémiologique».**

A partir de ces trois relations, est-il possible de préciser les différences entre parole et écriture dans la hiérarchie sémiotique ?

On pourrait considérer que la parole et l'écriture sont deux instanciations du même système, celui de la langue : tandis que la parole utiliserait des signifiants vocaux, les phonèmes, le second utiliserait des signifiants graphiques, les signes. La langue serait ainsi susceptible de s'incarner de deux façons différentes, l'une vocale et l'autre graphique. La manière graphique, venue plus tardivement tant du point de vue de l'espèce que du point de vue de l'individu, apparaît à première vue comme un système second qui serait issu de la parole. Mais en fait, il n'est pas possible de caractériser l'écriture comme un système issu de la parole, car leur relation n'est pas constituée sur le modèle des trois relations mises au jour par Benveniste. Du point de vue de l'engendrement en effet, il n'y a d'engendrement qu'entre systèmes sémiotiques de même nature (par exemple, entre l'alphabet normal et l'alphabet Braille). La parole et l'écriture ne font pas partie du même système sémiotique : la parole n'engendre pas

---

[447] **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 60.
[448] Cf. **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "Sémiologie de la langue", p. 54.

l'écriture.

Du point de vue de l'interprétance, il n'est pas absurde de se demander si ce n'est pas l'écriture qui permet de conscientiser la pratique de la parole et non l'inverse. En effet, la possibilité de constituer une grammaire passe par la maîtrise de l'écriture qui permet de réfléchir consciemment ce qui n'est que schème plus ou moins conscient[449] : tant qu'elle reste parlée, la langue n'est pas encore constituée en système grammatical, c'est-à-dire en système régi par des *règles*. C'est l'écriture qui permet de dégager de la pratique de la parole ses règles grammaticales.

C'est donc par la seule relation d'homologie que les systèmes sont comparables. De ce point de vue, les correspondances entre la parole et l'écriture existent, tout en étant limitées. Il y a de multiples signes écrits qui ne correspondent à rien dans la parole (par exemple, la ponctuation, mais aussi toutes les lettres qui ne sont plus prononcées et qui ne marquent plus que l'étymologie), de même que, inversement, l'intonation ne se lit pas dans l'écriture. C'est pourquoi le système graphique et le système de la parole semblent posséder des rapports de dépendance spécifique par rapport à la langue et qu'il n'est pas possible de considérer l'écriture comme une simple transcription de la parole. La spécificité des deux systèmes incarnant la langue se manifeste dans leur rapport au discours. Alors que le caractère évanescent de la parole implique immédiatement une situation pragmatique prise en compte par le discours, cette situation pragmatique semble ne pas avoir besoin d'être prise en compte dans le cas de l'écriture. C'est particulièrement vrai dans le cas de l'écriture imprimée, cas qui exige un traitement particulier.

14. Différence entre écriture manuscrite et écriture imprimée

On remarque que ce qui distingue l'écriture manuscrite de l'écriture imprimée est que cette dernière généralise par la mécanisation des usages graphiques qui seraient autrement restés à l'état embryonnaire[450].

---

[449] Cf à ce sujet l'avant-propos au livre de **J. Goody**, *La raison graphique; la domestication de la pensée sauvage*, Éditions de Minuit, Paris, 1979.

[450] Goody voit dans l'invention de l'imprimerie un facteur important de la généralisation de l'écriture sans y voir un objet spécifique par rapport à l'écriture manuelle. Cf. **J. Goody**, *La raison graphique; la domestication de la pensée sauvage*, op. cit., p. 54.

L'exemple de l'usage de la notion de tableau est, de ce point de vue, particulièrement éclairant[451]. Il est possible de faire des tableaux à la main. Mais l'imprimerie a permis de généraliser l'usage des tableaux, à tel point que le classement par tableaux, ayant une certaine utilité du point de vue du contrôle social, a fini par engendrer un certain nombre d'objets qui n'étaient pas auparavant susceptibles d'êtres dénombrés, ainsi qu'un certain nombre de lois, de nature statistique, déterminant les objets en question. Ian Hacking remarque à ce propos [452]:

> **«La science newtonienne n'avait aucun besoin des probabilités, sauf comme d'un outil pour localiser les causes sous-jacentes. Les lois statistiques qui semblaient être des faits bruts et irréductibles furent tout d'abord découvertes dans les occupations humaines mais elles ne purent être remarquées qu'après que les phénomènes sociaux eurent été énumérés, tabulés et rendu publics. Ce rôle fut rempli par l'avalanche de nombres imprimés au début du dix-neuvième siècle».**

Historiquement, ce n'est sans doute pas un hasard que ce soit Babbage qui, au XIXème siècle, fit le projet de réunir les trois académies de Londres, Paris et Berlin en vue de rassembler en tableaux toutes les constantes de la nature et qu'avant lui, ce soit Leibniz qui conçut un projet similaire[453]. Leibniz et Babbage, les grands ancêtres de Turing pour ce qui est de la conception de la mécanisation du calcul, virent donc dans la tabulation généralisée le moyen de concevoir le monde comme relevant du système sémiotique de l'imprimé : ainsi l'écriture imprimée apparaît-elle comme le “milieu” que l'esprit et le monde auraient en commun. La réflexion de Turing, de ce point de vue, marquerait une nouvelle étape de la mise au jour de cette relation, lui pour qui «mécanisme et écriture sont […] presque synonymes»[454]. Cette nouvelle étape paraît être clairement décrite dans “Computing Machinery and Intelligence”.

En effet, si l'on se place du point de vue de la différence des sexes, dont on se rappelle qu'elle est le critère du jeu de l'imitation, on remarque qu'il existe une différence capitale entre l'écriture manuscrite et l'écriture imprimée : alors que l'écriture manuscrite laisse encore transparaître un trait physique du locuteur qui peut

---

451 Cf. **J. Goody**, *La raison graphique; la domestication de la pensée sauvage*, op. cit., chap. 4.
452 **I. Hacking**, *The Taming of Chance*, Cambridge University Press, Cambridge, 1990, p. 3.
453 Cf. **I. Hacking**, *The Taming of Chance*, Cambridge University Press, Cambridge, 1990, p. 57-58 pour Babbage et p. 18 pour Leibniz.
454 **A. M. Turing**, “Computing Machinery and Intelligence”, op. cit., p. 456.

mettre sur la piste de la distinction selon le sexe par la différence dans la graphie et doit, de ce point de vue, être apparentée au système de la parole qui, elle aussi, laisse apparaître cette distinction, l'écriture imprimée tend à *supprimer* tout trait physique particulier. C'est pourquoi la question des rapports de l'écriture et de la parole, qui nous a servi de point de départ, rebondit : ce ne serait pas la parole qui s'opposerait à l'écriture, puisqu'on doit ranger du même côté la parole et l'écriture manuscrite, mais ce serait l'écriture *manuscrite* qui s'opposerait à l'écriture *imprimée*. Que faut-il entendre exactement par la notion d'imprimé ? Il ne faut pas prendre ce terme dans sa signification empirique. La notion d'imprimé réfère non seulement au cas de l'écriture manuscrite transcrite par le moyen de l'imprimerie mais aussi, plus profondément, au cas de l'écriture manuscrite d'un langage symbolique, qu'il soit mathématique ou logique. Dans ce cas en effet, le langage symbolique ne transcrit pas une parole : un symbole mathématique ou logique est dépourvu d'interprétation phonétique, même s'il est susceptible d'être nommé[455]. Or c'est bien ce dernier cas qui nous intéresse ici au premier chef. Il faut essayer d'en tirer les conséquences sur la façon de concevoir le rapport de l'écriture imprimée au discours.

Si l'on suit le point de vue de Turing en effet, on doit conclure que l'usage de l'écriture imprimée - au sens particulier que nous donnons à ce terme - a des conséquences fondamentales sur la façon de concevoir l'aspect sémantique du discours. De même que l'écriture manuscrite et l'écriture imprimée n'avaient pas le même rapport à la langue, de même elles ne s'adresseraient pas de la même manière à l'instance du discours : alors que l'écriture manuscrite référerait encore l'aspect sémantique du discours à la situation *pragmatique et sexuée* de parole, l'écriture imprimée, de par l'aspect impersonnel de sa typographie, ne référerait pas l'aspect sémantique du discours à la situation de parole. C'est précisément ce point qui, du point de vue linguistique, ferait l'objet du jeu de l'imitation.

2. Discours, langue et écriture dans le jeu de l'imitation

La situation du jeu de l'imitation est, à l'évidence, une situation de discours :

---

[455] Par exemple, le symbole $\int$ n'a aucun rapport avec le signifiant linguistique [*intégrale*] pas plus que $\sqrt{}$ ne renvoie à [*racine*].

l'interrogateur pose des questions aux joueurs et leur communication possède la forme canonique de tout discours, celle du dialogue[456].

21. L'enjeu d'une partie

L'enjeu d'une partie est, du point de vue de l'interrogateur, de parvenir à se représenter les joueurs (c'est-à-dire à identifier leur sexe), malgré le fait que les règles du jeu ont éliminé la présence de tous les systèmes sémiotiques se rapportant au corps (intonations, gestes, écriture graphique, etc) et ont eu pour effet d'isoler un seul système de communication, celui de l'écriture imprimée[457]. Or on aurait pu croire que la situation mise en place par le jeu aurait permis de conserver la signifiance de la langue mais pas la signifiance du discours. Le discours serait alors considéré comme lié à la situation particulière de la prise de parole, situation de nature corporelle extra-linguistique, dans laquelle d'autres systèmes de communication interviennent. Du point de vue linguistique, la question que pose le jeu est donc la suivante : la signifiance du discours est-elle liée à la situation corporelle de l'énonciation ou pas ?

Pour Turing, le jeu tend à montrer que, du point de vue de l'interrogateur, *la situation de discours est préservée*, même si toute communication non linguistique véhiculée par les autres systèmes sémiotiques est supprimée. Le jeu de l'imitation conserverait donc la signifiance des signes *et* celle de l'énonciation, malgré l'absence de tout système sémiotique extra-linguistique. La signifiance du discours n'est donc pas liée, pour Turing, à la présence d'autres systèmes sémiotiques non linguistiques. La situation de discours doit donc être rapportée à la langue. Dans le cas du jeu de l'imitation, il s'agit de la langue en tant qu'elle est imprimée. Il y a cependant, dans ce cas, une difficulté. La langue qui sert de moyen de communication entre l'interrogateur et les joueurs reproduit une langue naturelle : dans les dialogues écrits par Turing, c'est l'anglais qui est employé, mais toute autre langue naturelle ferait l'affaire. L'écriture imprimée est donc le médium dans lequel le discours s'incarne : en tant que simple médium, il doit reproduire fidèlement le discours. Mais dans le jeu n°2, l'ordinateur qui répond aux questions *fournit* à l'interrogateur des réponses en anglais mais ne les

---

[456] **E. Benveniste**, *Problèmes de linguistique générale*, tome 2, "L'appareil formel de l'énonciation", p. 85.
[457] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 434.

formule pas ainsi, puisqu'il opère au moyen d'un langage logique. C'est après compilation et traduction dans des niveaux de langue successifs qu'il peut apparaître sous la forme imprimée d'une langue naturelle dans les réponses qu'il fournit à l'interrogateur. Il y a là un changement capital puisque l'écriture imprimée apparaît, selon le jeu n°1 ou le jeu n°2, soit comme la trace de la présence cachée de la langue naturelle, soit comme de celle d'un langage logique.

Si, comme Turing le soutient, l'interrogateur n'aperçoit pas de changement entre un joueur humain et un joueur-machine, *c'est que le langage logique est suffisant pour assumer sans perte la double signifiance de la langue naturelle*, par les signes et par l'énonciation. Pour Turing qui suppose une issue positive au jeu de l'imitation, il est donc possible de rapporter la signifiance du discours à l'ordre linguistique, si l'on entend par ordre linguistique non pas celui de la langue naturelle mais celui du langage logique. Ce changement de nature sémiotique passe par *l'abandon de la langue naturelle comme norme première* au profit du langage logique. Il faut, pour prendre la mesure de ce changement, en venir aux buts poursuivis par Turing dans le jeu.

22. Le but visé par Turing

Le but de Turing est de montrer que le langage logique peut remplacer le système de la langue et donc *se situer en position dominante à la tête de la hiérarchie des systèmes sémiotiques*. Il y aurait ainsi un système sémiotique absolument indépendant de la langue et plus général qu'elle. Dans ce cas, la langue n'est pas le seul interprétant des autres systèmes sémiotiques puisque la relation que Benveniste appelle *S -> L,* qui rapporte à la langue l'explicitation des autres systèmes sémiotiques[458], c'est-à-dire qui en est l'interprétant général, est incluse comme partie dans une nouvelle relation hiérarchique *S -> L -> L'*, où *L'* désigne le langage logique.

Il faut donc réussir à montrer qu'il est possible, par l'expérience du jeu de l'imitation, de se détacher du système de la langue naturelle grâce à l'expérience du passage du jeu n°1 au jeu n°2. Le jeu de l'imitation permettrait d'envisager une situation absolument inédite d'un point de vue linguistique : alors que dans la langue conçue comme le système fondamental de la hiérarchie sémiotique, il faut distinguer

---

[458] Cf. Supra, § 13.

radicalement le niveau sémiotique et le niveau sémantique, ce ne serait plus le cas avec le langage logique quand on l'interprète d'un point de vue nominaliste. Dans ce cas, il y a une *influence* du sémiotique sur le sémantique puisque le simple fait d'utiliser des symboles logiques est susceptible d'imiter - en fait, de se substituer à - la signifiance du discours. Ainsi dans le jeu, la fonction dévolue à l'imitation est-elle précisément d'opérer cette synthèse des deux niveaux, sémiotique et sémantique. Cette attitude vis-à-vis de la sémantique du discours aurait pour conséquence d'abolir la différence des sexes : le niveau sémantique serait directement impersonnel, détaché de toute situation pragmatique et de ce fait, immédiatement universel.

Cette "influence", typiquement nominaliste, peut-elle se justifier quand on sait qu'il existe une différence importante dans le jeu de l'imitation entre le but visé par Turing (fonder un système sémiotique qui soit autonome par rapport à la langue et par rapport à la différence sexuelle qu'elle manifeste) et le but atteint par lui (fonder un système sémiotique dont le rapport à la langue vise à écarter la différence sexuelle sans l'abolir) ?

23. Le but atteint par Turing

Le langage logique dans la mesure même où le but visé par Turing dans le jeu de l'imitation et le but atteint par lui n'est pas le même, ne doit pas avoir la place que Turing lui assigne dans son rapport à la langue naturelle. Il faut, pour déterminer son rôle, revenir au rôle joué par la peau dans l'élaboration linguistique du modèle computationnel.

La peau nous est apparue comme un *matériau physique signifiant* qui fondait la différence sexuelle, c'est-à-dire une différence à la fois physique et psychique. En tant que matériau *physique*, elle se situe hors de la langue et son statut est d'ordre *sémantique* : elle est un objet du monde différent de la langue. En tant que matériau *signifiant*, elle rend possible l'opposition entre signifiant et signifié nécessaire à la constitution de la langue et son statut est d'ordre *sémiotique*. Il y a donc bien une articulation du sémantique et du sémiotique quand on se réfère à la notion de peau. Où situer, dans cette optique, le cas particulier du langage logique ? Deux remarques doivent être faites à ce propos pour situer la place du langage logique dans son rapport à la langue naturelle.

La peau, comme on l'a vu, joue une rôle fondamental dans la constitution des images qui rendent possible la mise en rapport du niveau matériel du cerveau et du niveau abstrait du concept logique de machine de Turing[459].

D'une part, le langage logique ne semble pas constituer un cas particulier dans la mesure où il entre directement dans le cadre plus général de la langue.

D'autre part, comme on l'a vu aussi[460], l'usage du langage logique implique une contrainte sémantique spécifique qui porte sur le discret : les états de l'esprit pertinents pour le modèle computationnel sont des états discrets. C'est cette contrainte émanant de l'usage du langage logique qui impliquait de distinguer trois niveaux de description différents dans l'élaboration des représentations au sein du modèle computationnel - le niveau intuitif du sens, le niveau formel du langage logique et le niveau physique de l'effectuation - tout en remarquant la parenté entre le niveau intuitif et celui du physique. Cette parenté apparaît clairement ici : la contrainte sémantique qui touche le niveau du mental a rapport avec le niveau physique puisque Turing remarque qu'il faut *faire comme si* la matière était discrète[461]. Il existe donc une *double contrainte sémantique* imposée par l'usage du langage logique dans le modèle computationnel, l'une du point de vue mental qui implique de limiter les états pris en considération aux états discrets, et l'autre du point de vue physique qui implique de ne considérer la matière physique que comme un matériau discret.

Mais cette contrainte physique a ceci de particulier qu'elle n'est pas arbitraire puisqu'elle peut s'incarner physiquement dans un ordinateur. Le niveau de description du fonctionnement physique d'un ordinateur relève en effet de l'opposition binaire entre états considérés comme discrets : par exemple, pour un ordinateur qui fonctionnerait avec du courant électrique, on ne prend pas en compte les états intermédiaires entre le passage du courant et l'absence de courant, comme le fait

---

[459] Cf. Deuxième partie, chapitre 2.

[460] Cf. Cette partie, chapitre 3. 1, § 121. 3..

[461] **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 439 : «Tout se meut continûment. Mais il y a beaucoup d'espèces de machines que l'on peut *considérer* comme des machines à états discrets».

remarquer Turing[462]. Cette opposition physique binaire doit suffire pour rendre compte de toutes les différences sémiotiques du vocabulaire du langage logique. Aussi le langage logique tel qu'il est physiquement incarné dans un ordinateur se laisse-t-il interpréter autant du point de vue sémiotique de la langue que du point de vue sémantique du discours, précisément parce que les contraintes sémantiques qu'il impose sont en même temps des contraintes physiques.

La spécificité du langage logique tel qu'il est physiquement incarné dans un ordinateur est donc, par rapport à la langue naturelle, la suivante : d'une part il est, comme elle, le produit d'une articulation entre le sémantique et le sémiotique qu'il faut rapporter à la peau; d'autre part, il est susceptible d'incarner physiquement l'articulation du discours et de la langue parce que la contrainte sémantique qu'il impose aux états mentaux en les limitant à ceux que l'on peut discrétiser est physiquement justifiable. C'est précisément ce deuxième point qui le distingue de la langue naturelle telle qu'elle est parlée, dans la mesure où l'on ignore quelle contrainte physique émanant du cerveau pèse sur elle ou même s'il s'agit d'une contrainte différente de celle qui pèse sur le langage logique. Le modèle computationnel de l'esprit est légitimé par le fait que la contrainte sémantique émanant de l'usage des symboles logiques est justifié à la fois du point de vue de l'esprit et de la matière. C'est cette cohérence interne du modère qui lui assure sa portée.

Il y a cependant un point qui n'est pas justifié dans l'interprétation nominaliste du modèle : on accepte comme un fait le caractère discret imposé par la contrainte sémantique sans qu'on puisse en rendre raison. C'est pourquoi cette contrainte sémantique apparaît, tant qu'on ne la rapporte pas à la contrainte naturelle de la peau, comme une *abstraction* conceptuelle contraignant comme de l'extérieur la structure des états de l'esprit et de la matière[463]. Comment justifier le fait que cette abstraction s'applique au cas du cerveau ? Il faut essayer de voir s'il n'existe pas des contraintes *naturelles* qui seraient sous-jacentes à la contrainte nominaliste propre au langage

---

462 Cf. **A. M. Turing**, "Computing Machinery and Intelligence", op. cit., p. 439.

463 Comme le fait remarquer A. Hodges, le modèle computationnel se présente sous l'aspect suivant : «la thèse de Turing consiste à dire que le modèle de la machine à états discrets est la description pertinente d'un des aspects du monde matériel - à savoir la façon dont fonctionne les cerveaux». **A. Hodges**, *Alan Turing and The Turing Machine* in [*The Universal Turing machine*, Herken R. ed., Oxford Science Publications, Oxford University Press, 1988], p. 11.

symbolique et qui pourraient, en éclaircissant son origine, remédier à son aspect arbitraire[464].

---

[464] C'est ce que remarque J. Petitot en explicitant le point de vue de R. Thom : «Ce […] point de vue […] a pour intérêt principal de permettre de *contraindre* les structures linguistiques par des contraintes imposées par la structure de la réalité et des Gestalten perceptives et de faire l'économie d'hypothèses innéistes (par exemple de type chomskien) dans l'explication des universaux du langage. C'est celui adopté par René Thom». **J. Petitot**, *Morphogenèse du Sens*, I, Presses Universitaires de France, Paris, 1985, p. 40.

<u>Chapitre III</u>

Les contraintes naturelles
et l'origine du langage logique

On vient de voir que la contrainte discrète qui pèse sur le langage logique est une abstraction conceptuelle. L'aspect discret de cette contrainte apparaît, dans le cadre du modèle computationnel de l'esprit, comme un *fait* qui n'appelle pas de justification. Il faut, pour tenter d'en donner une, réussir à rendre compte du processus de constitution de cette abstraction. C'est d'ailleurs la voie suivie par Turing lui-même puisqu'il la décrit dans “Computing Machinery and Intelligence” comme relevant de l'induction scientifique. On exposera ici sa description du processus en question et on montrera ensuite en quel sens sa démarche peut être considérée comme aporétique. Il faudra alors essayer de justifier autrement le processus de constitution de l'abstraction en essayant de faire intervenir des contraintes naturelles sous-jacentes à la contrainte discrète du langage logique. Comme le fait remarquer J. Largeault [465]:

> **«Si on définit l'être humain par la raison, la logique, à cause de ses affinités avec l'identité, acquiert une grande importance. Si on définit l'être humain par la vie, la physique ou la biologie gagnent le premier rang d'importance. Mais il n'est pas exclu, en dépit des apparences, que la logique soit une production du vivant. Le jour où nous saurons si, à quelque niveau, le vivant n'engendre pas quelque ordre analogue à celui de la logique ou de l'arithmétique récursive, nous serons mieux fixés sur sa signification réelle. Actuellement nous n'avons que la spéculation et nos partis pris».**

---

465 **J. Largeault**, Préface au livre de Bernard Ruyer, *Logique*, P.U.F, Paris, 1990, p. 13.

La spéculation que nous faisons ici est donc de considérer la logique comme une production du vivant si l'on considère le niveau spécifique de la peau, interprétée comme ce qui susceptible d'opérer des contraintes naturelles sur le langage.

1. L'aporie de la constitution de l'abstraction dans "Computing Machinery and Intelligence"

Turing décrit la constitution de l'abstraction par le biais de la notion d'apprentissage. Quelle forme doit prendre cet apprentissage ? Turing fait remarquer qu'il ne serait pas possible à la machine d'apprendre par le biais d'un canal émotionnel. Que peut être un "canal émotionnel" pour une machine ? Par émotion, Turing entend l'action physique portée sur le corps en vue d'attribuer punition et récompense. Aussi, dans le cas de la machine, le canal émotionnel de la machine est un canal susceptible de recevoir punitions et récompenses sous la forme de coups et d'absence de coups. Que serait l'équivalent d'un "coup" et d'une "absence de coups" pour une machine ? Turing fait remarquer à ce propos [466]:

> **«La machine doit être construite de telle sorte que les événements qui précédent directement l'occurrence du signal de punition ont peu de chance d'être répétés, tandis que le signal de récompense accroît la probabilité de la répétition des événements qui y ont conduit. Ces définitions ne présupposent aucune sentiment de la part de la machine».**

Et pourtant, malgré cette absence de sentiments, c'est-à-dire malgré cette réduction dans le langage de communication à une opposition de deux signifiants qui peut être physiquement incarné, ce type de construction semble à Turing inadéquat : pour lui, la quantité d'informations apprise par ce biais ne peut pas dépasser le nombre total de récompenses et de punitions accordées, c'est-à-dire de coups reçus. Bref, un système de communication qui soit à la fois binaire (c'est-à-dire susceptible d'être "appris" par la machine) et émotionnel (c'est-à-dire qui

---

[466] **A. M. Turing**, "Computing Machinery and Intelligence", p. 457.

s'apparente à ce qui relève du corps) ne peut pas engendrer un véritable apprentissage. Il faut réussir à construire un canal de communication qui permette l'induction scientifique et par ce biais, l'accumulation des connaissances. Pour ce faire, on doit, dit Turing, envisager la constitution de canaux de communication qui soient entièrement non-émotionnels et avec eux, la constitution d'un langage entièrement symbolique, *dont l'origine n'est pas précisée* [467]:

> **«D'une autre façon, on pourrait avoir un système complet d'inférence logique construit à l'intérieur. Note : ou plutôt, "programmé à l'intérieur", car notre machine-enfant sera programmée sur un ordinateur digital. Mais le système logique n'aura pas à être appris»**

Pourquoi le système logique n'aurait-t-il pas à être appris ? Il faut supposer que, pour Turing, ce système est inné. Autrement dit, il n'avance aucune raison qui pourrait l'avoir amené à considérer que l'induction scientifique doive remplacer l'émotion (et le corps) pour parvenir à une accumulation des connaissances. C'est en effet cette absence d'origine qui permet à Turing de considérer comme possible le renversement de la hiérarchie sémiotique et le remplacement de la langue orale par une autre langue, non-émotionnelle, programmable et permettant l'induction.

Peut-on décrire cette origine ? Le problème revient, dans l'optique de l'apprentissage, à justifier la possibilité d'une induction pour la machine. Turing revient, pour ce faire, à la distinction entre l'intérieur et l'extérieur [468]:

> **«Les processus d'inférence utilisés par la machine n'ont pas besoin d'être tels qu'ils puissent satisfaire les logiciens les plus exigeants. Il pourrait par exemple ne pas y avoir de hiérarchie des types. Mais cela ne veut pas dire que des erreurs de type vont survenir, pas plus que nous sommes contraints de tomber de falaises sans barrières. Des commandements adéquats (exprimés *à l'intérieur* des systèmes, ne faisant pas parties des règles *du* système) tels que "N'utilise pas de classe sauf si c'est une sous-classe d'une classe qui a déjà été mentionnée par le professeur" peut avoir un effet similaire à celui de "Ne t'approche pas trop près du bord"».**

Il est donc nécessaire, pour qu'une induction ait lieu, que des commandements soient donnés de l'extérieur à la machine. Il y a donc deux types

---

[467] **A. M. Turing**, "Computing Machinery and Intelligence", p. 457.
[468] **A. M. Turing**, "Computing Machinery and Intelligence", p. 458.

de règles [469]:

> **«Certaines peuvent être “données d’autorité”, mais les autres peuvent être produites par la machine elle-même, par exemple par induction scientifique».**

On retrouve donc ici, d’un point de vue logique, ce qui nous était apparu comme le résultat le plus fondamental du jeu de l’imitation, à savoir l’absence de fondement réel de la différence entre le corps et l’esprit. C’est pour réussir à rendre compte de la genèse concomitante du corps et de l’esprit qu’il faut faire intervenir des contraintes naturelles qui interviennent dans la constitution d’un langage logique. Il faut, de ce point de vue, renoncer à l’hypothèse nominaliste qui accordait une efficacité intrinsèque aux symboles du langage dans la constitution des états mentaux.

2. Les contraintes naturelles dans le processus de constitution de l’abstraction

La difficulté liée à la détermination de ces contraintes naturelles vient de ce que l’on semble dissocier le langage et la pensée, cette dissociation donnant lieu à deux illusions symétriques, comme le fait remarquer Benveniste [470]:

> **«Il est de la nature du langage de prêter à deux illusions en sens opposé. Étant assimilable, consistant en un nombre toujours limité d’éléments, la langue donne l’impression de n’être qu’un des truchements possibles de la pensée, celle-ci libre, autarcique, individuelle, employant la langue comme instrument. En fait, essaie-t-on d’atteindre les cadres propres de la pensée, on ne ressaisit que les catégories de la langue. L’autre illusion est à l’inverse. Le fait que la langue est un ensemble ordonné, qu’elle révèle un plan, incite à chercher dans le système formel de la langue le décalque d’une “logique” qui serait inhérente à l’esprit, donc extérieure et antérieure à la langue. En fait, on ne construit ainsi que des naïvetés ou des tautologies».**

La constitution du modèle computationnel de l’esprit tombe-t-il sous le coup de cette critique ? Il faut, pour s’en assurer, étudier, les images employées par Turing dans “Computing Machinery and Intelligence”.

Il est important de souligner que, du point de vue psychologique, de la création du modèle computationnel de l’esprit, *le processus de constitution et*

---

469 **A. M. Turing**, “Computing Machinery and Intelligence”, p. 458.
470 **Benveniste**, *Problèmes de linguistique générale,* op. cit., t.1, p. 73.

d’apprentissage de la machine est aussi le processus de constitution et d’apprentissage de l’esprit. Aussi la question de la constitution du plan de l’abstraction passe-t-elle par celle de la naissance concomitante de la machine et de l’esprit. C’est à la description des étapes de constitution de cette “machine-esprit” que nous allons nous attacher maintenant.

Trois images sont utilisées pour décrire la notion de machine-esprit dans sa naissance et son développement : l’image de l’équipe d’ingénieurs, celle de la peau de l’oignon et celle de la machine-enfant. Seule l’image de la peau d’oignon vise à décrire explicitement le rapport entre les notions de machine et celle d’esprit, puisque l’image de l’équipe d’ingénieurs sert à décrire la constitution de la machine et celle de la machine-enfant son développement.

Dans l’image de l’équipe d’ingénieurs, la machine est tout d’abord caractérisée comme sans peau : elle n’a rien à voir avec la peau, qui n’ajouterait rien, dit Turing, à sa constitution[471].

Dans l’image de la peau de l’oignon, il faut arracher des couches de peaux successives pour essayer d’atteindre l’esprit. Puis, dans la même image, la machine est rapportée à la surface d’une peau terminale sous laquelle il n’y aurait plus rien[472].

On sait par ailleurs que la peau, dans l’image de l’apprentissage de la machine-enfant, est constituée en page d’écriture, puisque la constitution de la machine était identifiée à l’écriture d’un signe sur la peau[473].

C’est en raison de ce mouvement que Turing parvient, de façon apparemment paradoxale, à caractériser la machine comme à la fois non-discrète et mécanique, alors que le mécanique avait précisément été défini par son caractère discret. Essayons d’expliciter ces images successives.

21. La première image : l’équipe d’ingénieurs

---

[471] «On pourrait par exemple insister sur le fait que l’équipe d’ingénieurs devrait être toute du même sexe, mais ce ne serait pas vraiment satisfaisant, car il est probablement possible de construire un individu complet à partir d’une seule cellule, disons de la peau d’un homme.» **A. M. Turing**, “Computing Machinery and Intelligence”, p. 435-436.

[472] **A. M. Turing**, “Computing Machinery and Intelligence”, p. 455-456.

[473] **A. M. Turing**, “Computing Machinery and Intelligence”, p. p. 456.

On se rappelle que Turing se pose à lui-même l'objection suivante [474]:

> **«On pourrait par exemple insister sur le fait que l'équipe d'ingénieurs devrait être toute du même sexe, mais ce ne serait pas vraiment satisfaisant, car il est probablement possible de construire un individu complet à partir d'une seule cellule, disons de la peau d'un homme».**

La machine doit donc être constituée sans le secours ni d'un engendrement sexué ni de son substitut minimal, la peau. Selon Turing, elle doit donc être conçue "hors-peau". Nous allons, pour expliciter la signification de cette image, partir d'une série de remarques de R. Thom touchant les contraintes naturelles sous-jacentes aux contraintes linguistiques.

211. L'analogie de R. Thom entre l'embryologie et la linguistique

R. Thom, inversant le sens de l'expression de "programme génétique" qui, de façon nominaliste, compare le génome à un texte écrit part au contraire du principe suivant [475]:

> **«La Linguistique s'explique par une extension des mécanismes de la Génétique et non l'inverse».**

R. Thom montre alors qu'il est possible d'instaurer un parallèle entre le développement embryologique et la constitution d'un domaine proprement linguistique[476]:

> **«On se rappelle que le développement embryologique d'un Vertébré peut être décrit par un graphe divergent, dont le sommet représente l'œuf activé, et les branches les spécialisations correspondant aux trois feuillets principaux : Ectoderme, Mésoderme, Endoderme. Le développement se poursuit ensuite par ramification de ces trois branches … […] Or dans la théorie linguistique connue sous le nom de Grammaire Générative, on décrit ainsi la genèse d'une phrase parlée; soit une phrase transitive simple telle que : *le chat mange la souris*. La théorie lui associe un graphe en forme d'arbre, dont le sommet P constitue l'origine, et d'où se ramifient des branches aboutissant aux "symboles terminaux", les mots. il y a donc isomorphisme réel entre ces deux type de graphes, isomorphisme qu'on peut poursuivre sur le pan de la signification fonctionnelle des catégories grammaticales et des spécialisations cellulaires : l'ectoderme correspond grosso modo au sujet, qu'il limite spatialement (et dans l'espace des stimuli, par l'activité nerveuse); le mésoderme correspond au verbe, tissu spécialisé dans le déplacement spatial; l'endoderme, enfin, correspond à la proie, l'objet grammatical».**

---

[474] **A. M. Turing**, "Computing Machinery and Intelligence", p. 435-436.

[475] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, Bourgois, Paris, 1980, p. 151.

[476] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 150.

Pour essayer de rendre compte du parallèle entre les mécanismes biologiques et les mécanismes linguistiques, Thom fait remarquer qu'il faut se pencher sur l'évolution qui va de l'animal à l'être humain. Il demande donc à son lecteur de faire une expérience de pensée particulière qui consiste à envisager le psychisme animal de "l'intérieur"[477]. Thom prend ensuite comme relation primordiale du psychisme animal la relation de prédation et essaye d'analyser à partir de cette relation la façon dont l'animal parvient à se déplacer dans l'espace [478]:

«On peut penser que la vie psychique de l'animal est constamment capturée par certains automatismes […] liés à la perception des objets importants biologiquement tels que proies et prédateurs. En fait, une analyse des mécanismes mis en œuvre en Embryologie, amène à penser que, d'une manière qui n'est qu'à peine symbolique, le prédateur "est" sa proie, il s'identifie psychiquement à elle. C'est seulement lorsqu'il aperçoit une proie réelles, extérieure, que se produit la catastrophe de perception. Il redevient lui-même en même temps que se déclenche le processus moteur de capture de la proie. De ce point de vue, le symbolisme primitif lié à la prédation s'explique aisément : si la prédateur (B) s'identifie à sa proie (A), on comprend aisément qu'un "indice" de A soit *ispo facto* un indice de B, sans avoir à faire intervenir une propriété de transitivité de la relation : indice de.»

Ainsi la relation métonymique de type "indice de" est-elle dès l'origine sous-tendue par une relation métaphorique de complète fusion dans l'objet. On retrouve, comme on l'a vu chez Cassirer, cette double relation dans la constitution du langage humain. Thom en déduit ensuite que l'"ego" de l'animal n'est pas une entité permanente, parce que la fascination exercée sur lui par les choses tend à lui enlever toute possibilité d'une distance par rapport à celles-ci. Dès lors, la distinction sujet-objet n'existe pas non plus de manière permanente et l'animal n'est pas capable de constituer une représentation intérieure de l'espace qui aurait la forme de l'espace euclidien [479]:

«[…] l'espace [de l'animal] est une réunion de cartes distinctes, chacune associée à un ego bien défini et affectée à un comportement moteur ou physiologique bien défini (territoires pour chasser, pour dormir, pour l'accouplement, pour la nichée, etc …) et l'on passe d'une carte à l'autre par des repères spatiaux (visuels ou olfactifs) bien définis».

On va voir que ces remarques de R. Thom peuvent nous aider à préciser la

---

[477] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 273.
[478] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 274.
[479] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 274.

signification de l'image de l'équipe d'ingénieurs, si l'on interprète bien celle-ci comme une description de la création de la "machine-esprit".

212. Application au cas de la première étape de la genèse de la machine-esprit

Je me permettrai d'interpréter les remarques de R. Thom dans un sens psychologique, c'est-à-dire en considérant que c'est moins l'étude du psychisme animal opérée «de l'intérieur» qu'il est parvenu à décrire, que les étapes d'un processus qui articule le plan sémantique du discours et le plan sémiotique de la langue, processus propre au psychisme *humain*[480].

On se rappelle que, pour Thom, le proto-espace animal est constitué de niches séparées dans lesquelles l'animal se projette et par lesquelles il est capable de s'identifier totalement à l'objet. Dès lors, la "fascination" exercée par les choses sur l'animal fait bien de lui une entité "hors peau", puisque se réalise alors une fusion avec l'objet désiré et que la peau ne joue pas le rôle d'une limite entre l'intérieur et l'extérieur. De ce point de vue, le cas décrit par Thom de l'identification de l'animal à sa proie est identique au cas de la constitution d'une métaphore dans le domaine linguistique, dans la mesure où la métaphore est précisément le lieu où les sens de deux mots sans rapport se trouvent confondus.

C'est pourquoi l'image de l'équipe d'ingénieurs qui vise à constituer la machine comme un objet "sans peau" peut être considérée comme une métaphore. Aussi la constitution psychologique de la machine-esprit en tant qu'entité à laquelle il est possible de penser renvoie-t-elle au moment le plus primitif, "animal" dirait Thom, de la constitution de l'esprit humain par le biais du modèle de la machine. On retrouve ici la façon dont W. R. Bion parvient à décrire la genèse des pensées en opposant ce qui relève des contenants et ce qui relève des contenus de pensée. Bion décrit en effet le développement des contenus de pensée en faisant remarquer [481]:

---

[480] C'est d'ailleurs ce à quoi aboutissait R. Thom quand il précisait à la fin du même chapitre : «Il n'est pas interdit de penser que - selon la loi de récapitulation - l'enfant passe par une période d'aliénation primitive, où certains êtres et objets exercent sur lui une totale fascination». **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p.. 275.

[481] **W. R. Bion**, *Learning from Experience*, Basic Books, New York, 1962; trad. franç. sous le titre *Aux Sources de l'expérience*, P.U.F, Paris, 1979, p. 112-113.

> **«Le modèle pour la croissance des [contenus] est un milieu où sont suspendus les "contenus". On doit concevoir les contenus comme émergeant d'une base inconnue. Ainsi l'apparition d'une notion servant de *contenu* à la pensée est-elle conçue comme émergeant de *nulle part*.».**

L'opposition des contenants et des contenus de pensée correspond très directement à la façon dont Thom aborde la question de l'espace chez l'animal. De façon plus générale, on peut donc dire que cette première étape tente de décrire la genèse dans l'esprit de contenus de pensée non encore verbalisés.

Cette première étape de la constitution de la machine peut donc se caractériser par trois traits négatifs : elle n'est pas *sexuée*, elle n'est pas située dans *l'espace* et elle n'est pas soumise à la discursivité du *langage*. Pour ce qui est du premier trait, on se rappelle que le but de Turing était de soustraire au domaine du sexuel la constitution de la machine. L'objection de l'équipe d'ingénieurs avait en effet pour but (du moins, pour but avoué) d'éliminer la différence sexuelle et, ce faisant, de supprimer tout ce qui a trait au substrat physique particulier propre à l'espèce humaine. Pour ce qui est du deuxième trait, on vient de voir que le fait d'être "sans peau" empêchait toute localisation spatiale de la machine, puisqu'en tant qu'elle est "sans peau", elle n'est ni dans un espace intérieur ni dans un espace extérieur. C'est pourquoi enfin elle semble émerger de nulle part, sans qu'il soit possible de lui attribuer par le langage un lieu qui la localiserait, en tant que concept, dans un espace abstrait.

22. La deuxième image : la peau de l'oignon

Turing compare la recherche d'un esprit non-mécanique à un oignon qu'il faudrait dépiauter jusqu'à une peau terminale qui ne contiendrait rien[482]:

> **«L'analogie de la peau de l'oignon est aussi utile. En considérant les fonctions de l'esprit ou du cerveau, nous trouvons certaines opérations que nous pouvons expliquer en termes purement mécaniques. Nous disons que cela ne correspond pas à l'esprit véritable : c'est une sorte de peau que nous devons arracher si nous voulons trouver l'esprit véritable. Mais dans ce qui reste, nous trouvons une autre peau à arracher et ainsi de suite. En procédant de la sorte, arrivons-nous jamais à l'esprit "véritable" ou parvenons-nous finalement à la peau qui ne contient rien ? Dans ce dernier cas, tout l'esprit est mécanique (ce ne serait pas une machine à états discrets, cependant. Nous en avons discuté.)».**

---

[482] **A. M. Turing**, "Computing Machinery and Intelligence", p. 455-456.

Une série de remarques de R. Thom peut nous servir à expliciter la signification de cette image.

221. La constitution d'un espace abstrait par le biais des règles du langage

Thom fait remarquer, après avoir noté que la représentation globale de l'espace a pour support le corps propre, que la capacité humaine à constituer l'espace en totalité est associée à l'apparition du langage, interprétée comme l'un des facteurs fondamentaux de la libération par rapport à la fascination exercée par les choses ayant, biologiquement, un intérêt. Le langage permettrait non seulement la localisation des concepts dans un espace abstrait ainsi que leur combinaison possible mais aussi la constitution de l'espace lui-même. La constitution de l'espace en totalité apparaît donc de façon concomitante à la constitution d'un plan sémantique du discours caractérisé, lui aussi, par son aspect global. Pour R. Thom en effet, l'être humain, contrairement à l'animal, est capable de raccorder les "cartes" qui morcellent l'espace et de constituer ainsi, grâce à la possibilité d'une répétition indéfinie, une représentation globale d'un espace telle qu'elle a cours dans la géométrie euclidienne. Il fait, à ce sujet, référence à "l'intuition du continu"[483]:

> **«Alors, comment expliquer que les mathématiques puissent représenter le réel ? La réponse, je crois, nous est offerte par l'intuition du continu. L'itération indéfinie peut parfois conduire à un objet immédiatement saisissable comme infini en acte. C'est là, je pense, l'interprétation qu'il faut donner au paradoxe éléate d'Achille et de la Tortue. Le paradoxe permet de donner un sens concret à la somme infinie $1/2 + 1/4 + 1/8 + \dots + 1/2^n + \dots$ à savoir la distance concrète à laquelle Achille dépasse la tortue. Le continu géométrique donne ainsi la possibilité de donner un sens à des êtres exigeant une infinité d'opérations (comme un nombre réel doté de toutes ses décimales)».**

Cette "intuition" du continu est le fait de la mise au jour de la règle linguistique *finie* qui suffit pour en rendre raison, comme Thom l'indique lui-même dans son exemple du nombre réel doté de toutes ses décimales, c'est-à-dire d'un nombre réel calculable. C'est par le langage que le continu devient dicible. L'aspect infini de l'itération de la règle suffit donc, dans certains cas, à représenter

---

483 **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 292.

le continu dans la mesure où toutes les étapes récursives de l'itération n'ont pas besoin d'être effectuées dans le temps. Ainsi peut-on se faire une idée du continu par le biais de l'itération indéfinie de règles finies de calcul. De ce point de vue, on peut réussir à se faire une idée de la globalité de l'espace, même si son homogénéité n'est pas complètement assurée, du fait de l'existence de nombres réels incalculables. On va voir que ces remarques peuvent nous aider à préciser le sens de la deuxième image utilisée par Turing pour décrire les rapports entre machine et esprit.

222. Application au cas de la deuxième étape de la genèse de la machine-esprit

Si l'on suit l'analyse de R. Thom, on peut dire que l'image de la peau de l'oignon décrit un raccordement progressif de l'espace par le biais de règles récursives.

D'un point de vue linguistique, l'image de la peau de l'oignon relève du trope de la métonymie puisqu'elle décrit un raccordement de l'espace.

D'un point de vue psychologique, la fusion avec l'objet qui rendait possible la constitution d'une métaphore dans la première image n'est plus possible ici dans le mesure où l'objet avec lequel fusionner, à savoir l'esprit lui-même, est "introuvable" : aussi faut-il pour tenter de le "trouver" raccorder mécaniquement toutes les parties morcelées de l'espace, la garantie de la globalité de l'espace étant apportée précisément par la constitution d'une peau terminale caractérisée à la fois comme surface globale, comme machine et comme esprit. Il devient alors possible d'en rester à l'aspect fini des règles, ou, comme dit Turing, de considérer que tout l'esprit est mécanique bien que l'on sache qu'il est aussi non-discret, comme l'a établi la métaphore de l'équipe d'ingénieurs. Mais cet aspect non-discret et métaphorique relève d'une étape antérieure dans la constitution psychologique de la "machine-esprit", celle dans laquelle machine et esprit sont à proprement parler confondus. Dans la deuxième étape au contraire, il existe une mise à distance relative des deux termes dans la mesure où leur rapport est de nature métonymique.

Les trois traits négatifs concernant le rapport à la sexualité, à l'espace et au langage reçoivent donc maintenant une détermination positive : la “machine-esprit” possède une détermination sexuelle minimale par le bais de la peau et elle est à la source de la constitution d'un espace au moyen de règles itératives de nature linguistique. Si l'on reprend le vocabulaire de W. R. Bion, on peut dire que le changement qui affecte les trois traits caractéristiques permet de constituer la machine-esprit en une enveloppe formant un *contenant* de pensée. Bion décrit le développement des contenants de pensée en utilisant l'image d'un réticule composé de parties accolées qu'il assimile à des manches de vêtements [484]:

> **«[Le contenant de pensée] se développe par apposition de manière à produire une série de “manches” qui sont conjointes. Il en résulte un réticule où les trous sont les manches […]».**

Il est donc possible de concevoir, à partir de l'image de l'enveloppe servant de contenant, la différence entre un espace intérieur et un espace extérieur, la peau jouant le rôle d'un bord entre l'intérieur et l'extérieur de l'espace. Dans la cas de la machine de Turing dont la fonction est précisément d'appliquer indéfiniment des règles d'itération, il faut essayer de saisir comment Turing justifie, au moyen de l'image de l'apprentissage de la machine-enfant, l'apparition de ces règles.

23. La troisième image : la machine-enfant

La comparaison de l'apprentissage d'une machine à celui d'un enfant, que nous avons déjà mentionnée, correspond à la troisième et dernière étape de la genèse de la “machine-esprit”. Elle a pour point de départ cette hypothèse émise par Turing selon laquelle «mécanisme et écriture sont […] presque synonymes» et que le cerveau de l'enfant comporte «peu de mécanisme et beaucoup de pages blanches»[485]. Comme précédemment, une remarque de R. Thom va permettre d'expliciter la signification de cette dernière image.

231. Le rôle du symbole écrit dans l'invention des concepts

---

[484] **W. R. Bion**, *Aux Sources de l'expérience*, op. cit., pp. 112-113.
[485] **A. M. Turing**, “Computing Machinery and Intelligence”, p. 456.

On avait vu que c'est à partir de l'écriture imprimée que, dans le montage expérimental du jeu de l'imitation, Turing avait pu croire pouvoir effacer entièrement la différence des sexes et constituer par ce biais ce que l'image de la machine-enfant appelle un «canal de communication non-émotionnel». R. Thom remarque que c'est par le biais de l'écriture - et il faut entendre ici écriture imprimée au sens que nous donnons à ce terme - qu'il devient possible de se libérer de la "fascination" exercée par les objets sur le psychisme[486] :

> **«En explorant une nouvelle théorie, en "bricolant" sur ce nouveau matériel, le mathématicien peut parfois percevoir une expression A - ou une relation - qui revient sous sa plume avec une insistance gênante; il sera alors tenté d'introduire un nouveau symbole pour condenser cette expression en une seule, et reprendre ainsi le calcul sur de nouvelles bases. Cette simple procédure peut parfois conduire au succès; plus souvent, il aura l'idée des expressions nouvelles à condenser, des nouvelles figures à construire et à nommer en suspectant *a priori* les propriétés. Introduire un nouveau symbole, c'est, en jetant une lettre sur le papier, […] libérer la démarche mentale des présences obsessionnelles qui l'entravent».**

Ainsi l'écriture imprimée jouerait-elle un rôle capital dans la genèse des nouveaux concepts du fait que le symbole matériel est, comme le faisait remarquer Cavaillès, «condition de création par sa mobilité dans le sensible»[487]. *Il y aurait bien alors une générativité propre au symbole* qui proviendrait de son aspect matériel et qui en ferait un nouveau support pour l'intuition quand les processus psychologiques de fusion et d'identification à l'objet reçoivent une contrepartie linguistique (métaphorique et métonymique) et permettent, de ce fait, la constitution d'une représentation de l'objet en question. C'est cette générativité objective qui aurait, dans le cas du modèle de la "machine-esprit", l'écriture imprimée pour support, comme l'indique la troisième image utilisée par Turing.

232. Application au cas de la troisième étape de la genèse de la machine-esprit

On se rappelle que l'interprétation psychologique du jeu de l'imitation

---

486 **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 276.

487 Cavaillès dans *Méthode axiomatique et formalisme; essai sur le fondement des mathématiques*, Hermann, Paris, 1938, réédition 1981, p. 94 cité dans **H. Sinaceur**, *Jean Cavaillès, philosophie mathématique*, Presses Universitaires de France, Paris, 1994, p. 62.

avait montré que le but recherché par Turing - montrer la possibilité d'une abolition de la différence des sexes par le biais de la création de la notion de "machine-esprit" - n'était pas atteint. On peut maintenant préciser pourquoi : il y a bien une générativité *objective* de l'écriture imprimée, comme l'ont remarqué Turing et R. Thom, mais celle-ci n'abolit pas l'aspect subjectif *aussi* présent dans cette générativité : c'est précisément ce qui apparaît dans cette troisième image, celle de l'écriture d'un signe sur la peau. Cette dernière étape de la création de la notion de "machine-esprit" apparaît alors comme le lieu de l'articulation d'une générativité objective du symbole et d'une générativité subjective de la pensée. Ainsi c'est sur cette générativité objective sur laquelle viendrait se greffer la méditation subjective propre à Turing sur la différence des sexes dans le processus de création : c'est pourquoi la "fascination" exercée sur lui par certains objets n'est pas définitivement écartée par le moyen de l'écriture imprimée mais seulement *déplacée* et se manifeste dans l'expression par le biais d'images qui, en renvoyant à la peau, décrivent la constitution progressive d'un expression linguistique où s'opère la synthèse des éléments objectifs et subjectifs du processus créateur.

Ainsi la création du modèle computationnel de l'esprit comme genèse concomitante de la notion de machine et de celle d'esprit peut-elle être rapportée à la possibilité *générale* de le décrire par le langage imprimé. Cette genèse met progressivement en rapport l'espace continu et le langage discret par l'intermédiaire d'une réflexion sur la peau dans sa signification psycho-biologique. On peut donc dire que la description de la genèse du modèle de la "machine-esprit" ne réduit pas l'esprit au mécanique mais que le mécanique lui-même n'a de sens que par rapport à une psychologie *dont les schèmes relèvent du vivant*.

Quelle place faut-il alors accorder au langage dans ce modèle ? Faut-il, en reconnaissant une certaine efficacité intrinsèque aux symboles, admettre l'interprétation nominaliste qui a présidé, chez Turing, à la constitution du modèle de la "machine-esprit" ?

3. La place du langage dans la genèse psychologique du modèle de la machine-esprit

On vient d'essayer d'analyser comment justifier l'hypothèse nominaliste qui accorde une efficacité aux symboles dans la constitution des états mentaux ou des états de la machine. Mais on a aussi montré en quel sens il fallait la compléter, dans la mesure où elle met essentiellement l'accent sur les aspects objectifs de la générativité propre aux symboles. Cette générativité objective ne peut avoir de sens que s'il est possible de montrer comment on peut subjectivement la *reconnaître* et lui accorder un sens. Comment le langage abstrait et objectif fait-il sens pour un sujet ? Les trois étapes que nous venons de décrire ont tenté de répondre à cette question en montrant comment la genèse d'un concept est aussi la genèse de sa description dans le langage. Les trois images utilisées par Turing dans “Computing Machinery and Intelligence” pour décrire la genèse concomitante de la machine et de l'esprit sont donc aussi la description de la façon dont le langage intervient dans cette genèse : à la fusion avec l'objet de prédation correspond la constitution du schème linguistique de la métaphore telle qu'il est décrit dans la première image, tandis qu'à la mise à distance de l'objet en un endroit localisé dans un espace abstrait correspond la constitution du schème linguistique de la métonymie telle qu'il est décrit par la deuxième image. La troisième image, en tant qu'elle met en rapport mécanisme et écriture imprimée, met en relief la spécificité de l'écriture imprimée par rapport à la parole et interprète l'écriture comme le moyen d'une articulation des données objectives et subjectives dans la création des concepts.

L'hypothèse nominaliste qui accordait une efficacité intrinsèque aux symboles était donc une hypothèse qui avait tendance à hypostasier l'aspect objectif de la générativité des symboles en écartant le cheminement subjectif vers l'abstraction, seul capable d'opérer une synthèse entre les aspects objectifs et subjectifs de la création du modèle computationnel de l'esprit.

Chapitre IV

Généralisation du modèle de Turing

Deux séries de remarques vont clore cette étude du modèle computationnel de l'esprit tel qu'il a été créé par Turing : la première est d'ordre historique et la seconde épistémologique. On remarque en effet, en étudiant dans une perspective historique les rapports des notions de machine et d'esprit, que non seulement Turing n'est pas le premier à avoir tenté une synthèse des deux notions mais qu'il n'est pas non plus le premier à l'avoir fait dans les termes qui sont les siens et qui font appel à des données à la fois objectives et subjectives. On s'appuiera pour le montrer sur deux exemples, l'un en “amont” de Turing et qui analysera le cas de Babbage et l'autre en “aval”, qui reposera sur une analyse de ce que R. Thom entend par “stratégie herméneutique”. C'est à partir de cette double confirmation historique de l'idée selon laquelle la constitution d'une notion passe par une synthèse linguistique d'aspects subjectifs et objectifs que nous tenterons de montrer la portée épistémologique du modèle de Turing.

1. Remarques historiques

11. En “amont” de Turing : Babbage

On sait que Turing se réclame, dans “Computing Machinery and Intelligence”, de l'autorité de Babbage[488].

Or Babbage avait écrit une autobiographie, *Passages from the Life of a Philosopher* [489], dont un chapitre porte le titre de “Jeux d'adresse”. Ce chapitre

[488] Babbage : né à Teignmouth en 1792, mort à Londres en 1871.

[489] **C. Babbage**, *Passages from the Life of a Philosopher*, London, Longman, 1864 republié dans [**P. Morrison** et **E. Morrison eds.**, *On the Principles and Developpment of the Calculator and*

me semble être une source majeure de “Computing Machinery and Intelligence”[490]: il permet d’expliquer non seulement le vocabulaire de Turing (les termes de machine, d’automate, de test, sont *tous* empruntés à Babbage) mais, dans une certaine mesure, l’idée même du jeu de l’imitation. Il paraît opportun, pour s’en rendre pleinement compte, de donner ici un large extrait de ce chapitre :

> **«Longtemps après que la traduction du mémoire de Menabrea[491] eut été publiée et après avoir réalisé beaucoup de dessins de la Machine Analytique et de ses parties, je commençai à réfléchir aux moyens intellectuels grâce auxquels j’étais parvenu à des résultats si avancés et si inattendus. Je repassai dans mon esprit les différents principes que j’avais rencontrés dans mes articles publiés et non publiés et je m’arrêtai avec satisfaction sur le pouvoir que je possédai sur le mécanisme [*mechanism*], grâce à l’aide de la Notation Mécanique[492]. Je sentis, cependant, qu’il serait plus satisfaisant pour l’esprit des autres, et même dans une certaine mesure pour le mien, d’essayer de faire l’épreuve du pouvoir de ces principes tels que je les avais conçus, en prenant pour objet des questions d’une espèce tout nouvelle que je tenterais de résoudre au moyen de ceux-ci, eux qui m’avaient guidés avec tant de succès dans d’autres cas.**
>
> **Après y avoir longuement réfléchi, je choisi pour mon test d’inventer une machine [“*machine*”] qui serait capable de jouer avec succès un jeu de nature purement intellectuelle, tel que le “morpion”, le jeu de dames, le jeu d’échecs, etc.**
>
> **Je m’efforçai de m’informer des opinions de personnes de toute condition et de tout âge, pour savoir si elles pensaient que le fait de jouer un jeu d’habileté requérait l’intervention de la raison humaine. La réponse à peu près constante était affirmative. Certains faisaient observer à l’appui de leur thèse que si ce n’était pas le cas, un automate [*automaton*] pourrait jouer à ces jeux. Un petit nombre de ceux qui étaient versés dans les sciences mathématiques laissait ouverte la possibilité d’une machine [*machinery*] capable d’une telle action; mais ils niaient tous de façon véhémente la possibilité de construire une telle machine, du fait des myriades de combinaisons que même le jeu le plus simple impliquerait.**
>
> **Au sujet de la première partie de mon enquête, je parvins bientôt à une démonstration que tout jeu d’adresse est susceptible d’être joué par un automate.**
>
> **Des considérations plus poussées montrèrent que, quelle que soit la position des jetons [*any position of the men*] sur l’échiquier (que cette position soit possible ou impossible), si l’automate pouvait effectuer adéquatement le premier coup, alors il [*he*] devait être capable de gagner le jeu, en supposant toujours que, compte tenu de la position des jetons, cette conclusion soit possible.**
>
> **[…]**
>
> **En conséquence, toute la question de faire qu’un automate joue à un jeu quelconque dépendait de la possibilité pour la machine de se représenter la totalité des myriades de combinaisons qui lui était liée. En autorisant cent coups pour chaque adversaire dans le**

---

other seminal writings, by Charles Babbage and others, Dover Publications, New-York, 1961], pp. 3 à 157.

490 Cela ne m’a malheureusement été ni confirmé ni infirmé par le biographe de Turing, A. Hodges, à qui j’ai posé la question.

491 Il s’agit d’un ingénieur italien qui fit une traduction française des plans de la machine analytique de Babbage. Son mémoire, publié à la Bibliothèque Universelle de Genève fut par la suite traduit et annoté par la comtesse Lovelace. Cf. [**P. Morrison** et **E. Morrison eds.**,*On the Principles and Developpment of the Calculator* (…), op. cit.], p. 225 sq.

492 La “notation mécanique” était une notation inventée par Babbage en vue de décrire les mouvements de la Machine Analytique. On trouve une description de cette notation dans son article : “Laws of Mechanical Notation”, reprint dans [**P. Morrison** et **E. Morrison eds.**, *On the Principles and Developpment of the Calculator* (…), op. cit.], pp. 357-362.

cas de la plus longue partie d'échecs, je trouvai que les combinaisons nécessaires au fonctionnement de la Machine Analytique dépassait énormément toutes celles qui étaient requises, même pour le jeu d'échecs.
[…]
La nouvelle difficulté consistait en ceci que, lorsque l'automate avait à se déplacer, il pouvait arriver qu'il y eut deux coups différents, chacun permettant de remporter le jeu. Dans ce cas, il n'y avait pas de raison dans la machine qui puisse diriger son [*his*] choix : à moins que des dispositions soient également prises, la machine tenterait d'exécuter deux mouvements contradictoires.
Le premier remède que je trouvai pour parer ce défaut fut de faire en sorte que la machine garde trace du nombre de jeux qu'elle [*it*] avait gagné depuis le début de son existence. Chaque fois que deux coups, que nous pouvons appeler A et B, conduisaient également à un jeu gagnant, l'automate était construit de telle sorte qu'il consultait la liste du nombre de jeux qu'il [*he*] avait gagné. S'il arrivait que ce nombre fût pair, il [*he*] était dirigé vers la solution A; si le nombre était impair, il [*he*] devait choisir la solution B.
S'il y avait trois coups également possible, l'automate était dirigé de telle sorte qu'il divise le nombre des jeux qu'il avait gagné par trois. […]
Il est clair que l'on pouvait résoudre ainsi un nombre quelconque de possibilités. Un spectateur curieux qui observerait les jeux joués par un automate pourrait observer longtemps, avant qu'il ne découvre le principe grâce auquel celui-ci agit. On doit aussi remarquer combien cela illustre les meilleures définitions du hasard données par le philosophe et le poète :

"Le hasard n'est que l'expression de l'ignorance de l'homme" -Laplace
"Tout hasard n'est qu'un dessein mal compris" -Pope

Me satisfaisant pleinement de ma capacité à construire un tel automate, l'étape suivante était de me demander s'il y avait une probabilité quelconque pour que l'automate puisse me rapporter, dans un temps modéré et si l'on en faisait la démonstration publique, une somme d'argent suffisante pour construire la Machine Analytique. Un ami auquel j'avais, à une époque antérieure, communiqué mon idée, avait entretenu de grands espoirs quant à son succès pécunier. Quand on apprendrait qu'un automate pouvait battre non seulement les enfants mais aussi papa et maman à un jeu d'enfants, il ne semblait pas déraisonnable de supposer que tout enfant qui entendrait parler de lui demanderait à sa maman de le voir. Réciproquement, toutes les mamans et quelques papas qui en entendraient parler emmèneraient sans aucun doute leurs enfants voir ce spectacle si singulier et si intéressant».

Ce texte de Babbage est remarquable à plus d'un titre.

On y voit tout d'abord apparaître le terme de *test*. Sa fonction est d'emporter l'adhésion des sceptiques, tel le Professeur Jefferson de "Computing Machinery and Intelligence". Turing n'use donc du terme de test que dans le contexte bien particulier d'un argument contre les sceptiques et en imitant la démarche même de Babbage. Le "test" de Turing n'est donc pas, comme nous l'avons montré par d'autres biais, un véritable dispositif expérimental qui aurait valeur démonstrative : c'est plutôt un argument *rhétorique* cherchant à convaincre un "spectateur" potentiel. Babbage introduit donc lui aussi, comme Turing le fera après lui, un spectateur extérieur au jeu : l'issue du jeu est censée favoriser l'adhésion à l'argument selon lequel la machine possède un caractère généralement attribué à l'être humain, à savoir la raison. Pourquoi ce "test" favoriserait-il l'adhésion du spectateur ?

Parce que, du point de vue du spectateur, l'automate capable de décider entre deux coups également gagnants se comporte comme un être humain qui fait des choix et non pas comme un être définitivement pris dans une contradiction indécidable, tel l'âne de Buridan. Il faut donc, pour emporter l'adhésion du spectateur, *occulter* le mécanisme de la division qui rend possible l'apparence d'un choix. La découverte par le spectateur de ce que le choix relève en fait d'un algorithme est liée au calcul des probabilités, comme chez Turing. La différence entre les deux jeux provient de ce que dans le jeu de l'imitation, le jeu se pratique "en aveugle"[493] et que l'interrogateur ne voit pas les autres joueurs, contrairement au jeu de Babbage dans lequel le spectateur voit la machine jouer.

La conséquence la plus importante de l'occultation du mécanisme du choix est qu'elle modifie le *genre* de la machine. Il est en effet remarquable de constater que Babbage mélange les genres grammaticaux et décrit au masculin [*he* et *his* au lieu de *it* et *its*] un automate qui, selon toutes les règles de la grammaire anglaise, ne peut être décrit qu'au neutre[494].

A quelle règle obéit cette transformation non-grammaticale ? Non pas au vocabulaire employé : que ce soit le terme de machine [*machine*], de "machinerie" [*machinery*] ou d'automate [*automaton*], la confusion des genres s'effectue de la même manière. C'est plutôt au critère du choix que cette transformation obéit : chaque fois que la machine doit faire un choix, Babbage utilise le masculin. Aussi peut-on dire que la machine devient masculine quand elle *apparaît,* pour le spectateur, comme opérant un choix : l'occultation qui se manifeste par le calcul des probabilités interprété comme manque de connaissances, se traduit donc également sur le terrain grammatical, qui renvoie

---

493 On sait que, chez Turing, ce caractère "aveugle" concerne non seulement la vue mais aussi le toucher.

494 Heims dans **S. J. Heims**, *John von Neumann and Norbert Wiener; From Mathematics to the Technologies of Life and Death*, MIT Press, Cambridge, Massachussets, 1980, p. 83 avait noté "en passant" cette confusion des genres. Il avait seulement ajouté entre parenthèses "sic" dans la citation d'une partie de ce texte de Babbage, sans essayer de donner une explication de la nature de la confusion en question.

indirectement au domaine sexuel, par la transformation du neutre en masculin. On voit donc ici la naissance même de ce qui deviendra chez Turing le jeu de l'imitation dans ses formules successives, jeu qui fera de l'occultation du sexe des participants la règle même de son fonctionnement.

On doit aussi relever, pour expliquer cette confusion des genres, un indice dans le texte lui-même, à savoir l'usage très particulier que Babbage fait de l'expression «n'importe quelle position des pions» : Babbage se sert, pour exprimer le terme de "pion" du terme utilisé uniquement pour le jeu de dames, dans lequel les pions peuvent être appelés les "hommes" [*men*]. Cette expression étant toujours employée au pluriel [*men*, *they*] par Babbage, on ne peut pas savoir si son usage au singulier l'aurait conduit à confondre le masculin et le neutre [*he* à la place de *it*] comme dans le cas de la machine, c'est-à-dire si la "faute" de syntaxe se serait étendue à cette expression. Cependant, tout se passe comme si le fait que les pions soient appelés les "hommes" induise, comme par contagion, que ceux qui manipulent les pions soient également des hommes, qu'ils soient en réalité des êtres humains ou des machines. C'est en effet *après* avoir utilisé cette expression que la confusion des genres dans la description de la machine apparaît dans le texte. Cette confusion n'est pas le fait du spectateur qui est autorisé à voir qu'il y a une différence entre les joueurs, être humain et machine. Elle est donc bien le fait de Babbage lui-même et de la position de "survol" qu'il occupe par rapport au jeu qu'il décrit.

Il y a donc, quand on se penche sur les raisons de la confusion "syntaxique" des genres, deux types de raison qui entrent en jeu : l'une, intérieure au jeu, qui porte sur la notion de choix (aux yeux d'un spectateur, la machine passe pour un homme grâce à ses "choix"), l'autre, extérieure au jeu, qui porte sur la façon dont Babbage a connaissance à la fois de ce que peut connaître le spectateur et de ce que lui, Babbage, connaît (pour Babbage s'identifiant au spectateur, les pions "masculins" sont déplacés par des hommes).

Résumons l'argumentation. Pour le spectateur, la machine fait un choix. Le spectateur attribue donc une humanité à la machine. Pour Babbage, cette attribution d'une humanité à la machine par le spectateur est vraie, puisque c'est lui, Babbage, qui a monté le mécanisme de division des coups déjà gagnés. Mais c'est précisément ce que le spectateur *ne doit pas* découvrir : il ne doit pas avoir la possibilité d'identifier l'auteur (masculin) du mécanisme en question. Ce que le spectateur ne doit pas découvrir, c'est donc précisément le fait qu'il ait *raison* en attribuant le mécanisme du choix à un homme. Bref, rendre la machine masculine renvoie au fait que Babbage se projette de l'extérieur (à partir de sa position omnisciente de survol) dans le spectateur (qui n'a qu'une perception tronquée) : c'est cette position de survol à l'intérieur et à l'extérieur du jeu, qui rend la détermination sexuelle de la machine possible.

Il est à peine besoin d'insister sur la parenté avec la problématique du jeu de l'imitation. D'autres indices vont d'ailleurs dans le même sens qui corroborent la parenté des démarches[495]. On doit souligner également l'aspect autobiographique du texte, aspect que l'on a retrouvé dans "Computing Machinery and Intelligence" : l'article de Turing peut facilement passer pour l'autobiographie de Turing, comme en écho à l'autobiographie de Babbage. Sans entrer dans les raisons qui ont pu pousser Babbage à commettre ces "erreurs" grammaticales et qui demanderaient une étude approfondie de l'œuvre et de la vie du grand prédécesseur de Turing, on peut remarquer seulement - et c'était le but

---

[495] On trouve également chez Babbage d'autres idées dont Turing semble avoir tiré profit dans "Computing Machinery and Intelligence". Mentionnons quatre d'entre elles. La première concerne le fait que la seule façon de battre un joueur expérimenté (et, dans le cas de Turing, quoi de plus expérimenté qu'une machine ?) est de *mal* jouer. Cf. **C. Babbage**, *Passages from the Life of a Philosopher*, London, Longman, 1864 republié dans [**P. Morrison** et **E. Morrison eds.**, *On the Principles and Developpment of the Calculator* (…), op. cit.], p. 29 : «Je remarquai dans ces occasions que si je jouais selon des ouvertures classiques, telles qu'on les trouve dans les livres, j'étais certain d'être battu. La seule façon que j'avais d'espérer gagner était de faire au début de la partie un coup si mauvais qu'il n'avait été mentionné nulle part dans un traité. Brande [l'adversaire de Babbage] possédait et avait lu à peu près tous les livres sur le sujet». La seconde idée, plus anecdotique, concerne l'un des dialogues fictifs du jeu de l'imitation dans lequel l'interrogateur demande de rédiger un poème. Babbage rapporte dans sa biographie qu'après avoir demandé à un ami poète la vitesse avec laquelle il parvenait à rédiger des vers, celui-ci lui avait répondu qu'il ne pouvait pas écrire plus de quatre vers par jour. (Cf. **C. Babbage**, *Passages from the Life of a Philosopher*, London, Longman, 1864 reprint dans [**P. Morrison** et **E. Morrison eds.**, *On the Principles and Developpment of the Calculator* (…), op. cit.], p. 77). C'est précisément cette impossibilité à composer de la poésie sur le champ que Turing prend comme exemple dans le dialogue fictif du jeu de l'imitation. La dernière idée provient du titre même de deux chapitres de l'autobiographie de Babbage : "Expérience avec l'eau" (chapitre XV) et "Expérience avec le feu" (chapitre XVI) dans lesquels Babbage décrit un certain nombre d'inventions techniques liées à la maîtrise de ces éléments. Il n'est pas impossible que ces deux chapitres aient évoqué pour Turing le poème de Felicia Hemans sur "Casabianca" qui décrit précisément une "expérience" liée à l'eau et au feu.

de ces remarques historiques sur Babbage - que l'attitude de Turing n'est pas une attitude isolée et qu'il existe, dans la construction qu'il a faite, des éléments généraux qui ne dépendent pas de sa subjectivité personnelle, mais qui relève bien de la subjectivité en général.

On peut en avoir aussi confirmation en étudiant un cas qui se situe historiquement en "aval" de Turing.

12. En "aval" de Turing : R. Thom

R. Thom a tenté d'analyser la notion de jeu grâce à la notion de "boîte noire". Par "boîte noire", il entend tout système dont l'interaction s'opère par le biais d'entrées et de sorties, sans que soit directement connue la façon dont s'opère la transformation des entrées en sorties. Le contenu d'une boîte noire, c'est-à-dire la transformation qu'elle opère, n'est donc pas représentable directement. Pour avoir une connaissance approchée de cette transformation, c'est-à-dire pour dévoiler la transformation interne opérée par une boîte noire, le seul moyen, dit Thom, «est de jouer avec»[496].

Deux stratégies de jeu sont concevables.

121. Les stratégies réductionnistes et herméneutiques

121. 1. La stratégie réductionniste

La stratégie la plus commune «consiste à casser la boîte noire pour voir ce qu'il y a dedans»[497]. Il s'agit d'une stratégie réductionniste répandue dans les disciplines expérimentales et qui rencontre des difficultés épistémologiques bien répertoriées.

---

[496] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 303.
[497] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 304.

On peut citer en deux : d'une part, casser une boîte noire pour en analyser ses composants suppose que les composants en question ont résisté à la destruction de la boîte et qu'il forment donc chacun en eux-mêmes une autre boîte noire plus difficile à casser; d'autre part, il faut supposer que l'on peut envisager entre les éléments d'une même boîte noire des rapports relativement stables.

121. 2. La stratégie herméneutique

Une autre stratégie, que Thom qualifie "d'herméneutique" consiste au contraire à ne pas chercher à casser la boîte noire pour en analyser le contenu, c'est-à-dire à ne pas tenter de se représenter la transformation opérée par la boîte noire à partir de l'analyse de ses éléments, mais à *s'identifier* à la boîte noire en question en lui prêtant notre intelligence [498]:

> **«[…] imaginons qu'un "esprit", un psychisme (*the ghost in the machine*) dirige au moins partiellement le système, et essayons de nous mettre "dans sa peau"».**

Il n'est pas besoin de souligner combien le choix du terme de peau est ici significatif, quand on garde à l'esprit la fonction fondamentale qu'elle revêt dans la constitution même du "jeu de l'imitation". Aussi peut-on ajouter que prêter une intelligence à une boîte noire pour réussir à en interpréter le comportement implique qu'on lui attribue une peau, attribution symbolique qui place la boîte noire, désormais animée, dans la sphère d'une interrogation touchant ce qui relève de la sexualité. Cette attribution symbolique n'est pas identique à une représentation telle que celle-ci était conçue dans la stratégie réductionniste, parce que l'attribution symbolique, en l'occurrence le don d'une peau, est à *l'interface* du physique et du mental, alors que la représentation se voulait être entièrement mentale.

Thom définit ensuite "la tâche hernéneutique" de la manière suivante [499]:

> **«La tâche herméneutique, en face d'une boîte noire particulièrement énigmatique pourra s'assimiler à un jeu, dont l'interprétant et "l'esprit dans la boîte" seront les**

---

[498] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 305.
[499] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 306.

joueurs».

Dans le contexte qui est le nôtre, si l'on se rappelle que le jeu de l'imitation lui-même était apparu comme une image idéalisée du corps, on peut interpréter ce corps comme formant une "boîte noire" particulièrement énigmatique. En quoi est-elle énigmatique ? Précisément en ce que l'attribution symbolique de la peau ne relève pas d'une opération seulement intellectuelle, comme dans le cas de ce qui relève d'une représentation, mais d'une *projection* à la fois physique et intellectuelle.

122. Les outils mathématiques des stratégies de jeu

122. 1. La fonction comme outil de la stratégie réductionniste

Dans le cas de la stratégie réductionniste, l'outil mathématique le plus simple à utiliser est, pour Thom, le concept de fonction. En effet, dans le cas des boîtes noires où l'entrée détermine complètement la sortie, c'est-à-dire où l'histoire de la boîte noire n'influence pas le résultat en sortie, la notion de fonction permet de «rechercher des phénomènes qui admettent une telle description»[500]. Il s'agit d'un type particulièrement simple de boîtes noires, puisque la sortie se laisse décomposer par le biais d'une fonction univoque de l'entrée et qu'il n'est donc pas nécessaire de lui supposer des «états internes»[501].

122. 2. Deux modèles de la stratégie herméneutique

D'autres "boîtes noires" plus compliquées sont imaginables. Elles

---

[500] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 303.
[501] **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 303.

nécessitent une stratégie de type herméneutique et partant, d'autres outils mathématiques que l'utilisation "simple" de la notion de fonction. R. Thom en invoque deux, à partir de la structure d'un jeu décrit par I. Ekeland qui n'est pas sans rappeler le jeu de l'imitation[502]:

> **«Chacun des deux joueurs, Pierre et Jean, a à sa disposition un espace qui lui est propre : P pour Pierre; J pour Jean. Jouer, c'est pour chaque joueur choisir un point dans son espace : p dans P; j dans J. Une fois ces choix faits (indépendamment l'un de l'autre), une troisième instance (la "banque") détermine les gains respectifs des deux joueurs G_p ; G_j comme fonctions G_j (p;j); G_p (p;j) des points choisis. Le but de chaque joueur est de maximiser son gain.**
>
> **Mais supposons alors que la présence de Pierre soit inconnue de Jean, Jean est alors en présence d'une "boîte noire", dont l'entrée est le point j dans J, et la sortie le gain G_j (p;j). On peut alors se poser le problème inverse : dans quel cas une boîte noire à sortie réelle peut-elle être associée à un "jeu" à deux joueurs ? ».**

Thom montre alors que si l'on dote l'esprit de la boîte noire en question d'une intentionnalité *permanente*, on doit utiliser la notion de fonction potentiel telle qu'elle a été définie au sein de la théorie des catastrophes élémentaires. Mais il est aussi possible de projeter une intentionnalité moins élaborée et il faut, dans ce cas, utiliser un modèle statistique. Dans ce dernier cas, R. Thom montre qu'à une entrée dans la boîte noire correspond un nuage de points en sortie, que l'on peut caractériser par l'existence d'un "démon" choisissant le point de son espace propre de manière stochastique, satisfaisant à la loi des grands nombres. Thom en conclut [503]:

> **«En ce cas, dans sa tentative d'interprétation, l'esprit humain n'a conféré à son partenaire qu'un psychisme très rudimentaire, celui du *matelot ivre* dont la démarche erratique engendre le mouvement brownien… Entre ce cas et l'intentionnalité permanente définie par un potentiel, il y a sans doute toute une classe de dynamiques intermédiaires à découvrir, placées entre le déterminisme têtu d'un potentiel, et la spontanéité gratuite du choix "arbitraire"; elles seront sans doute plus propres à simuler le comportement réel du psychisme humain».**

122. 3. Le cas "intermédiaire" du jeu de l'imitation

Le cas du jeu de l'imitation, dans la mesure où il sert de fondement à la description du modèle de la "machine-esprit" au moyen de l'outil mathématique qu'est la machine de Turing, nous paraît être un des cas intermédiaires permettant

---

502 **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 306.
503 **R. Thom,** *Modèles mathématiques de la morphogenèse*, 2ème édition, op. cit., p. 307.

de simuler le comportement du psychisme humain dont parle R. Thom.

En tant qu'elle rend précise la notion de fonction calculable, la notion de machine de Turing représente bien un cas intermédiaire parce qu'elle nécessite de faire appel à la notion d'état interne à la boîte noire. Cas-limite de la notion de fonction dans la mesure où elle associe déterminisme dans le processus de calcul et indéterminisme quant au résultat, la machine de Turing, si on la considère comme une boîte noire avec laquelle il est possible de jouer, se voit donc dotée d'une intentionnalité qui, sans être permanente, n'est pas non plus entièrement aléatoire.

Le cas du modèle de la "machine-esprit" apparaît ainsi comme susceptible d'entrer dans une typologie générale des "boîtes noires" une fois que l'on accepte de l'envisager d'un point de vue herméneutique, c'est-à-dire une fois que l'on a reconnu le rôle fondamental accordé à la fonction dissimulatrice dans la constitution de la notion de représentation. C'est à partir de considérations de ce type que l'on peut avancer quelques remarques épistémologiques sur la portée du modèle de Turing.

2. Remarques épistémologiques sur le modèle de Turing

La notion de machine de Turing a ceci de particulier que le dévoilement d'un aspect de la notion mathématique de fonction, son aspect calculable, semble exiger simultanément le dévoilement d'un aspect du psychisme, la psychologie du calculateur. Ce dévoilement n'est pas sans poser des problèmes épistémologiques qui sont liés à la façon dont on conçoit le rapport des mathématiques et de la psychologie. On peut en effet comprendre ce rapport comme une explicitation de la psychologie grâce à un concept mathématique - et l'on adopte alors une attitude réductionniste - ou, inversement, comme une mise au jour du rôle de la psychologie dans l'élaboration des concepts mathématiques - et il s'agit alors d'une attitude herméneutique - . Selon que l'on interprète le rapport en question dans le premier ou le second sens, on a tendance à privilégier l'aspect mathématique ou l'aspect psychologique du modèle. Or ces deux sens semblent coexister au sein du modèle en question. L'intérêt épistémologique de l'étude de la constitution du modèle computationnel vient donc de ce qu'elle exige de faire intervenir à la fois une attitude réductionniste et une attitude herméneutique et d'en comparer les apports.

Comme le fait remarquer R. Ruyer, on assimile généralement la notion de science à l'attitude réductionniste[504] :

> **«Il y a deux façons d'être informé : 1°. par *participation* à un autre "je", ou à une conscience spécifique, dans la mémoire, l'instinct, la formation instinctive; 2°. Ou par *observation* (d'un autre objet ou d'un autre comme objet) dans tous les autres cas. La science ne considère généralement que l'information par *observable*; elle méconnaît l'information par *participable*».**

C'est à cause de cette tendance générale à la connaissance par observation que l'interprétation du modèle de Turing a donné lieu jusqu'à présent à de multiples études dans lesquelles l'accent a été mis sur le pouvoir explicatif du modèle d'un point de vue réductionniste. Ces études voient dans le modèle de Turing le premier essai de psychologie computationnelle et font reposer leur analyse sur un usage de symboles cognitifs au sens que D. Widlöcher donne à ce terme[505]. Ces interprétations sont en partie justifiées par les textes de Turing lui-même ainsi que par l'attitude déclarée qui était la sienne et que nous connaissons par de nombreux témoignages. L'article le plus clair de ce point de vue est son article *princeps* de logique mathématique, que nous avons commenté dans la première partie, "On Computable Numbers …".

21. Connaissance par observation : le cas de "On Computable Numbers …"

On peut en effet interpréter l'attitude de Turing dans "On Computable Numbers …" comme une attitude réductionniste. Dans ce cas, la notion de machine de Turing peut servir à observer la psychologie du calculateur au travail. Si l'on reprend le vocabulaire de R. Thom, on peut décrire l'attitude de Turing dans son article en disant qu'il y a interprété l'esprit comme une boîte noire susceptible d'être cassée et étudiée dans ses composants primitifs, les *états*

---

[504] **R. Ruyer**, *Paradoxes de la conscience et limites de l'automatisme*, Albin Michel, Paris, 1966, p. 212.

[505] Cf. Introduction, deuxième partie.

mentaux discrets. Le résultat le plus patent, de ce point de vue, a été de montrer qu'en "cassant" cette boîte noire particulière qu'est l'esprit du calculateur, on n'obtient rien d'autre qu'une machine.

Pour parvenir à ce résultat, Turing a réussi à éviter le psychologisme. Au lieu de réfléchir à l'apparition psychologique d'un concept mathématique dans le psychisme considéré comme une boîte noire primitive, comme c'est par exemple le cas dans le livre du mathématicien J. Hadamard sur l'invention dans le domaine mathématique[506], Turing a montré que le concept mathématique de machine de Turing pouvait servir à "casser" la boîte noire de l'esprit, qui n'a plus alors statut de fondement primitif. Dans cette optique, on *applique* un concept mathématique à une réalité qui lui est extérieure, en l'occurrence la psychologie du mathématicien au travail. Mais cette application reste relativement arbitraire : le concept de machine de Turing peut se porter sur bien d'autres objets que la psychologie du mathématicien et il semble qu'il n'y ait aucune nécessité à envisager cet objet plutôt qu'un autre. On peut d'ailleurs montrer, comme Wang Hao l'a fait, que les considérations sur la psychologie du mathématicien peuvent être entièrement éliminées de la description des machines de Turing et être remplacées par des considérations objectives, par exemple sur le pavage du plan[507].

L'aspect relativement arbitraire de l'application de la notion de machine de Turing à la psychologie vient de la spécificité de l'objet qu'est l'esprit : en effet, en appliquant un concept mathématique à la psychologie, on occulte nécessairement la constitution *psychologique* du concept en question puisque, comme on l'a vu, la psychologie ne peut plus servir, dans ce cas, de fondement primitif. D'où proviennent alors les concepts mathématiques ?

Je considère que c'est le but profond de "Computing Machinery and

---

[506] Cf. **J. Hadamard**, *Essai sur la psychologie de l'invention dans le domaine mathématique*, (1945), traduction française, Gauthiers-Villars, Paris, 1959, p. 14 : «Dans cet ouvrage, j'utiliserai les résultats de l'introspection, les seuls dont je me sente qualifié pour parler. […] Les méthodes objectives - observations du dehors - sont celles où l'expérimentateur est différent du penseur. Obervation et pensée n'interfèrent pas; mais d'autre part, on n'obtient ainsi que des renseignements indirects dont on ne déchiffre pas facilement la signification».

[507] Cf. **W. Hao**, "Games, Logic and Computers", Scientific American, 1965, 213 (5), 98-106.

Intelligence", dont l'analyse a constitué la deuxième partie de notre étude, que d'essayer de répondre à cette question. Or la réponse implique d'adopter une attitude qui n'est plus d'observation mais de participation : c'est pourquoi le jeu de l'imitation a pu nous apparaître comme un cas particulier pouvant entrer dans l'esquisse de typologie générale des boîtes noires mise en place par R. Thom.

22. Connaissance par participation : le cas de "Computing Machinery and Intelligence"

Le but de "Computing Machinery and Intelligence" ne consiste pas à fonder objectivement la psychologie, contrairement à ce que laisse supposer une lecture superficielle du texte - et contrairement aussi à la tendance générale de l'attitude réductionniste adoptée en science -. C'est à la description de la naissance psychologique d'un concept mathématique que l'on assiste dans "Computing Machinery and Intelligence". La description de cette naissance ne privilégie ni l'attitude introspective d'un Hadamard, ni l'attitude objective du réductionnisme : l'article de Turing décrit la genèse *simultanée* d'un concept mathématique et du domaine psychologique que le concept éclaire, sans chercher à privilégier l'un des deux termes au détriment de l'autre. Il s'en suit que la notion de psychologie n'a plus du tout la même signification que celle que lui conférait l'attitude réductionniste : elle doit plutôt être considérée comme l'étude des expressions du corps en tant qu'il est le lieu de l'articulation de l'objectif et du subjectif et qu'il permet, à ce titre, de rendre compte de la genèse des concepts. C'est dans le cadre de cette psychologie qu'il nous paraît possible d'éclaircir la difficile question de l'invention dans les sciences.

23. Remarques sur l'invention

La description de l'invention du modèle en question doit apparaître comme la description *simultanée* du concept mathématique de machine et du domaine phénoménal de la psychologie du mathématicien. Comme nous avons essayé de le montrer, cette description passe par la mise au jour des contraintes naturelles qui ont présidé à sa constitution. Aussi souscrivons-nous aux remarques de G.

Simondon qui rapporte la capacité psychologique d'invention des êtres humains à leur existence biologique, parce que c'est le même schéma de finalité qui s'exprime dans les deux cas à des niveaux différents.

On sait que la survie de l'organisme biologique passe par une auto-régulation avec son milieu : un organisme vivant dans un milieu qui lui permet de perdurer dans le temps forme un système. Pour G. Simondon, c'est la représentation de cette auto-régulation d'un point de vue purement virtuel qui permet de rendre compte de l'invention de l'objet technique. Il y aurait de ce point de vue une auto-régulation virtuelle des concepts techniques dont l'invention serait liée à la constitution d'un "milieu associé".

Simondon prend l'exemple d'une turbine [508]:

> **«Tel est l'ensemble constitué par l'huile et l'eau en mouvement dans la turbine Guimbal et autour d'elle. Cet ensemble est concrétisé et individualisé par les échanges thermiques récurrents qui ont lieu avec lui : plus la turbine tourne vite, plus la génératrice dégage de chaleur par effet Joule et pertes magnétiques; mais plus la turbine tourne vite, plus la turbulence de l'huile autour du rotor et de l'eau autour du carter s'accroît, activant les échanges thermiques entre le rotor et l'eau. C'est ce milieu associé qui est la condition d'existence de l'objet technique inventé».**

Le "milieu associé" est donc le lieu de l'auto-régulation qu'il a fallu concevoir avant toute expérience puisque l'invention de l'objet technique a lieu *en même temps* que le milieu associé dans lequel il s'insère. L'invention technique reproduit de ce fait l'ordre du vivant :

> **«C'est parce que le vivant est un être individuel qui porte avec lui son milieu associé que le vivant peut inventer; cette capacité de se conditionner soi-même est au principe de la capacité de produire des objets qui se conditionnent eux-mêmes».**

C'est pour cette raison qu'il est possible d'établir un lien entre le fonctionnement de la pensée, l'invention d'un objet technique et la capacité qu'a la nature de transformer les formes vivantes. L'objet technique apparaît alors comme la condition de lui-même puisque sa réalisation consiste à associer dans un ensemble des éléments matériels qui étaient jusqu'alors épars. C'est cet auto-conditionnement de l'objet technique qui oblige, lors de son invention, à renverser

---

[508] **G. Simondon**, *Du mode d'existence des objets techniques,* op. cit., p. 57.

les rapports temporels et à le concevoir comme déjà existant au sein de son milieu associé [509]:

> **«Seule la pensée capable de prévision et d'imagination créatrice peut opérer ce conditionnement renversé dans le temps : les éléments qui matériellement constitueront l'objet technique, et qui sont séparés les uns des autres, sans milieu associé avant la constitution de l'objet technique, doivent être organisés les uns par rapport aux autres en fonction de la causalité circulaire qui existera lorsque l'objet sera constitué; il s'agit donc ici d'un conditionnement du présent par l'avenir, par ce qui n'est pas encore».**

Dès lors, l'aspect dynamique du fonctionnement de la pensée tel qu'il se manifeste dans l'invention de l'objet technique se constitue précisément dans la possibilité d'un renversement de l'ordre du temps. Aussi est-ce grâce à ce schème fondamental de l'objet technique, dans la phase qu'il constitue avec son milieu associé, qu'il est possible de justifier l'analogie entre l'invention technique et le fonctionnement de la pensée[510]:

> **«Le dynamisme de la pensée est le même que celui de l'objet technique; les schèmes mentaux réagissent les uns sur les autres pendant l'invention comme les divers dynamismes de l'objet technique réagiront les uns sur les autres dans le fonctionnement matériel».**

Les schèmes mentaux parviennent à rendre connexes des éléments auparavant épars. Ces éléments sont de tous ordres et il n'y a rien qui ne puisse servir à la constitution des schèmes. On a montré, dans "Computing Machinery and Intelligence", que le modèle computationnel de l'esprit s'est constitué à partir d'éléments épars - quelque fois même fossilisés depuis longtemps dans la mémoire de Turing - et qui ont fini par devenir connexes une fois qu'ils ont pu être rapportés au corps par l'intermédiaire d'une méditation sur la peau en tant qu'elle est le lieu de l'articulation linguistique des données objectives et subjectives. Le jeu de l'imitation n'est donc pas seulement une autobiographie de l'invention d'un concept mais il possède bien une portée *générale* en tant qu'il montre comment, chez un individu, s'articule par le langage des déterminations objectives et subjectives. De ce point de vue, les remarques historiques que nous avons faites en début de chapitre corroborent les remarques épistémologiques que

---

509 **G. Simondon**, *Du mode d'existence des objets techniques,* op. cit., p. 57.
510 **G. Simondon**, *Du mode d'existence des objets techniques,* op. cit., p. 58-59.

nous faisons à la fin.

Conclusion

Nous voudrions, au terme de notre travail, reprendre les articulations de cette étude pour tenter d'en mesurer la portée. L'intuition qui a guidé notre recherche et que nous avions empruntée à R. Thom[511] est la suivante : *une structure mathématico-logique discrète comme celle de la machine de Turing doit être plongée dans un continu pour avoir, psychologiquement, une signification*.

1. Le sens intrinsèque du concept mathématico-logique de machine de Turing

A l'évidence, la structure mathématique discrète du concept de machine de Turing possède un sens intrinsèque qui ne dépend pas directement de son rapport à la notion de continu. Il semble donc à première vue possible d'étudier la constitution de ce concept et de lui seul. On remarque de ce point de vue que l'intérêt épistémologique du concept de machine de Turing est double : il intervient lorsqu'on observe la naissance de la théorie de la calculabilité; il intervient encore s'agissant de la naissance de l'informatique et celle de l'intelligence artificielle. Cette double - ou peut-être même triple - naissance exigeait donc que l'on pose un certain nombre de questions d'ordre historique avant d'en venir à la description abstraite du concept de machine de Turing lui-même.

11. La naissance de la théorie de la calculabilité

Du point de vue de la naissance de la théorie de la calculabilité tout d'abord : comment la question du calcul finitaire est-elle venue au premier plan à partir de la constitution de la perspective métamathématique mise en place par

[511] Cf. Avant-propos, § 24.

Hilbert ? Quel rôle a joué, dans la critique adressée à Hilbert par Brouwer à propos du formalisme, la question de l'essence du calcul ? Où se situe, enfin, l'apport spécifique de Turing sur cette question ? En abordant ces points, nous nous sommes aperçu que la question de la détermination arithmétique du continu était constamment présente chez ces mathématiciens et qu'il était impossible de l'éliminer sous peine d'appauvrir notre enquête épistémologique. Il y avait là un premier indice qui corroborait la remarque de R. Thom sur la nature des concepts mathématico-logiques discrets dont la signification dépend de leur rapport au continu.

D'autres questions se sont donc ajoutées d'elles-mêmes aux premières : quel rôle la détermination arithmétique du continu a-t-elle joué dans le contexte mathématique du formalisme et de sa critique par Brouwer ? Quels rapports ces mathématiciens ont-ils eux-mêmes établi entre la maîtrise du continu et la nature de la pensée ? Dans le cas du concept de machine de Turing tout particulièrement, deux points nous ont paru remarquables touchant le rapport de la détermination conjointe de la nature du continu et de celle de la pensée.

Du point de vue de la maîtrise du continu tout d'abord, c'est bien par rapport à la question de la calculabilité des nombres réels formant le continu que le concept de machine de Turing - contrairement à d'autres concepts qui lui sont formellement équivalents - a été constitué : la structure mathématique discrète déployée par le concept de machine de Turing semblait donc n'avoir de sens qu'en tant qu'elle permettait de rendre compte d'un aspect de la nature du continu - son aspect calculable. Cette structure permettait d'envisager une classe très étendue de nombres réels tout en prouvant, négativement, l'existence de réels hors de cette classe. Il semblait donc bien que la question du continu ait eu un rôle moteur dans l'investigation menée par Turing.

Du point de vue de la détermination de la nature de la pensée ensuite, la question de la maîtrise du continu a posé directement, et ce, dès la mise en place de la perspective métamathématique de Hilbert, la question du pouvoir de la pensée ainsi que celle de son fonctionnement. En faisant du point de vue métamathématique une “cour de cassation” pour juger mécaniquement de la validité des théorèmes mathématiques, le tribunal ultime de la raison devenait nécessairement caractérisé par son aspect finitaire. Dès lors, la pensée apparaissait comme ayant le pouvoir de maîtriser finitairement le continu grâce à un fonctionnement “mécanique”. Aussi déterminer la nature précise de ce qu’il fallait entendre par fonctionnement mécanique devait-il avoir des implications immédiates sur la nature de ce qu’il fallait entendre par pensée. C’est un des aspects remarquables de l’enquête menée par Turing d’avoir réussi à proposer une version du concept de mécanique qui satisfasse à la fois l’exigence de la détermination du continu et sur celle de la pensée. Cette double détermination a eu des conséquences sur la naissance de l’informatique et sur celle de l’intelligence artificielle.

12. La naissance de l’informatique et celle de l’intelligence artificielle

Constitué à partir d’un certain nombre d’acquis théoriques et techniques datant de la seconde guerre mondiale, le projet britannique de construction d’un ordinateur programmable a paru quelque peu marginal par rapport au projet américain concurrent doté de moyens beaucoup plus considérables, bien qu’il eut été en avance sur ce dernier. Nous n’avons pas étudié les développements simultanés de ces deux projets mais seulement mis l’accent sur la forme que la participation de Turing y avait prise.

En tant qu’auteur de “On Computable Numbers …” et que membre du service britannique du chiffre, l’influence théorique des travaux de Turing a été considérable jusqu’à la fin des années quarante mais s’est estompée par la suite pour deux raisons au moins. D’une part, n’ayant pas reçu une formation en ingénierie qui lui aurait permis de suivre dans tous leurs détails les développements de la construction matérielle des composants de l’ordinateur, il s’est, dès le début des années cinquante, éloigné de la recherche dans ce qui était en train de devenir l’informatique. D’autre part, peu enclin par tempérament à se fondre dans un groupe de recherche composé d’un grand nombre de participants, il a préféré retrouver des terrains d'investigation qu'il pouvait explorer seul[512].

Dans les années cinquante, il a jeté les bases d'une nouvelle discipline qui s'appellera - plus de vingt ans après -, l'intelligence artificielle et qu'il a conçue, le premier, à partir du slogan : «construire un cerveau». Ainsi l'informatique théorique était-elle à peine constituée en objet scientifique que Turing concevait déjà ce qui, dans cette science, pouvait avoir une portée à la fois mathématique et psychologique. Ce nouveau projet ne doit pas nous étonner dans la mesure où il s'inscrit dans la continuité de la recherche que Turing avait menée sur les fonctions calculables et qui conduit à réunir, par le biais de ce qu'il est convenu d'appeler la "thèse de Turing", la question abstraite de la calculabilité et celle des conditions psychologiques d'accès à celle-ci.

Il devenait alors possible, une fois la constitution progressive du projet de Turing précisée, de décrire le concept de machine de Turing tel qu'il avait été exposé à l'origine par Turing.

13. La description du concept de machine de Turing

Après la description du concept lui-même, on pouvait envisager sa finalité du point de vue de la détermination du domaine des fonctions calculables et préciser ses limites en soulevant la question logique de la décidabilité par le biais du "problème de l'arrêt", devenu classique en informatique. On pouvait alors essayer de dresser un inventaire des outils théoriques mis en place par Turing - ils n'ont pas toujours eu de postérité immédiate, ni dans le vocabulaire ni dans les concepts de la théorie de la calculabilité ou de l'informatique - : les notions de machine circulaire et non-circulaire, celle de machine à oracle ainsi que la place qu'il convient d'accorder à la notion de machine universelle en sont les exemples privilégiés, dans la mesure où l'on peut retrouver, à l'occasion de l'exposition de ses outils théoriques, le fil directeur contenu dans la "thèse de Turing" qui consistait à mettre en rapport l'aspect objectif et l'aspect subjectif présent dans la

---

[512] Il faut noter de ce point de vue l'extraordinaire capacité de renouvellement intellectuel qui est la sienne et qui lui a permis de se plonger le tout premier dans l'étude de la morphogenèse par le biais de sa modélisation informatique.

notion de calcul.

L'aspect purement épistémologique de notre travail a donc permis de préciser l'itinéraire de Turing et ses motivations scientifiques. De ce point de vue, la question de la détermination de la nature du continu - mathématique et physique - a joué un rôle considérable parce que c'est par rapport à elle que le concept discret de machine de Turing s'est déterminé. Ainsi, même s'il était possible de décrire le concept de machine de Turing sans envisager ses liens avec ce que nous avons appelé “la question du continu”, il nous a paru indispensable d'évoquer cette question pour rendre justice au concept de machine de Turing lui-même. La remarque de R. Thom qui nous sert de fil directeur ne visait d'ailleurs pas à dire que les structures mathématiques discrètes n'avaient pas de sens mathématique si on ne les considérait pas dans leur rapport au continu mais que leur signification en tant qu'elles peuvent servir de *modèle*, exigeait qu'elles soient plongées dans le continu. C'est donc la signification du concept de machine de Turing en tant qu'il peut servir de modèle qu'il nous fallait, dès lors, étudier pour préciser ses liens avec la question du continu.

2. Le sens extrinsèque du concept de machine de Turing : sa signification en tant que modèle

Dans quelle mesure le concept de machine de Turing peut-il avoir, outre son sens mathématico-logique intrinsèque, la signification d'un modèle ? Si l'on suit la remarque de R. Thom, un concept ne deviendrait psychologiquement significatif et n'acquiererait de ce fait le statut de modèle que s'il est plongé dans un substrat continu, parce que c'est par ce biais qu'il acquiert une signification. Quelle serait la spécificité du concept de machine de Turing quand il est plongé dans un substrat continu ?

La spécificité de la transformation du concept en modèle une fois plongé dans le continu proviendrait du fait qu'il ne modélise pas n'importe quelle activité mais qu'il constitue un modèle *de la psychologie*, c'est-à-dire qu'il permet de faire réflexion sur la constitution des significations. Il s'agirait donc d'un méta-modèle qui, en modélisant l'activité du cerveau, selon l'expression de Turing, parviendrait du même coup à rendre raison des étapes de la constitution des significations. C'est en ce sens que le modèle issu du concept de machine de Turing peut servir à fonder une théorie des représentations.

21. L'intelligence artificielle comme modèle de l'esprit

Le projet de la "construction d'un cerveau" s'est heurté dès l'origine à un certain nombre de préjugés qui, sans avoir gêné Turing dans sa recherche proprement dite, lui ont paru devoir être combattus parce qu'ils représentaient une sorte d'obstacle "sociologique" à la diffusion d'idées nouvelles portant sur la nature et le fonctionnement de la pensée. C'est pour mener ce combat que Turing a été conduit à écrire l'article philosophique si curieux de "Computing Machinery and Intelligence" que nous avons longuement commenté.

Cet article, qui a joué par la suite le rôle d'une "charte" pour l'intelligence artificielle, a suscité de nombreuses interprétations qu'il fallait examiner. Nous avons essayé de montrer combien les interprétations logicistes et fonctionnalistes étaient "sélectives" et laissaient de côté un certain nombre de propositions, généralement considérées comme ironiques ou contingentes et qui, de ce fait, étaient censées ne pas mériter que l'on s'attarde à leur donner du sens. Notre principe de conduite a été exactement inverse : tenter d'attribuer un sens à ce qui pouvait paraître incongru, même s'il fallait aller le chercher dans l'itinéraire personnel de Turing, parce que la constitution d'un concept en modèle est tout d'abord une entreprise individuelle de recherche du sens, cette recherche passant *aussi* par des données personnelles. Cela paraissait tout particulièrement nécessaire dans le cas d'un modèle de la psychologie parce qu'il y a, dans ce cas, une quasi-nécessité à faire une place à ce qui relève de l'individuel. Ce n'est que dans un deuxième temps qu'il devient possible de généraliser le sens du modèle si celui-ci le permet, c'est-à-dire s'il est possible de montrer comment il est arrimé à des contraintes naturelles qui lui confèrent une portée générale. L'investigation touchant ces contraintes naturelles nous a conduit à mettre l'accent sur l'importance du domaine biologique dans la constitution des modèles et tout particulièrement du modèle de l'intelligence artificielle.

22. L’origine biologique des modèles : le cas de l’intelligence artificielle

Que le projet d’une intelligence artificielle ait été formulé au moment précis où Turing entreprenait l’étude informatique de la morphogenèse ne nous paraît pas indifférent : la référence à la question de la nature du *vivant* occupait l’esprit de Turing quand il a jeté les premières bases du projet d’intelligence artificielle. La remarque que nous empruntions à R. Thom en commençant prend dès lors toute sa portée : la question de la *signification* du concept de machine de Turing a surgi sous la forme de la possibilité d’une *intelligence* de la part de la machine - d’une “intelligence artificielle” - quand Turing s’est interrogé sur la portée du concept en question pour les êtres vivants que nous sommes. Il nous semble, de ce point de vue, que c’est dans la mesure où la structure mathématico-logique discrète de la machine de Turing était plongée dans ce que Turing appelle le «continu matériel» au sein duquel évoluent les organismes, qu’il a pu concevoir le projet d’une intelligence artificielle. Ainsi la question du vivant et celle du sens sont-elles intimement liées si l’on se place du point de vue du continu. C’est ce que nous avons tenté de montrer dans notre analyse de “Computing Machinery and Intelligence”.

Aussi est-ce un paradoxe du point de vue de l’histoire des sciences de voir comment la figure de Turing a pu être réappropriée à titre de père fondateur par ceux qui, à la fin des années soixante, ont conçu un projet d’intelligence artificielle fonctionnaliste qui séparait radicalement le niveau computationnel de “l’esprit” du niveau des substrats matériels. A lire attentivement Turing en effet, on se rend compte qu’il y a des substrats qui ne nous sont pas indifférents en tant qu’êtres vivants parce que c’est par leur biais que nous nous maintenons en vie. A ce titre, notre corps est bien le substrat fondamental à partir duquel des significations deviennent possibles, y compris celles qui sont exprimées par l’intelligence artificielle.

Dès lors, le modèle de Turing pour la psychologie ne pouvait pas servir tel quel de fondement pour une psychologie computationnelle de nature cognitive, à moins de prendre en compte la dimension du corps incarné et la façon dont s’exprime dans le langage cette dimension. Aussi le modèle de Turing nous apparaît-il finalement plus propice à constituer une théorie philosophique de la représentation qu’à servir de fondement scientifique à la psychologie computationnelle.

23. Conséquences philosophiques

Nous nous sommes placé, dans notre étude, d’un point de vue *non-dualiste* et nous avons tenté d’envisager les conséquences qu’impliquait un tel point de vue pour l’interprétation du concept de machine de Turing. Une fois trouvé l’angle d’attaque que représente ce que nous avons appelé la “question du continu”, les deux directions prises par notre travail - analyse du concept et analyse d’un modèle - se sont alors imposées à nous. Il peut paraître curieux d’avoir mené une enquête sur le concept de machine de Turing dans des directions au premier abord aussi étrangères que sa description épistémologique d’une part et les conditions naturelles de son existence en tant que modèle d’autre part, quand on sait que la description de ces conditions naturelles nous a entraîné sur le terrain de la psychologie particulière d’un individu. S’il est vrai qu’il n’y a pas de science de l’individuel, alors que viennent faire les remarques que nous avons faites concernant la psychologie de l’individu Turing ?

Remarquons tout d’abord, qu’on ignore habituellement tout de la vie psychique des scientifiques et que l’histoire d’un concept ou d’une théorie se trouve de ce fait coupée de sa base psychologique. Il se trouve que dans le cas de Turing, grâce au remarquable travail de reconstitution mené par A. Hodges ainsi qu’aux nombreux témoignages de personnes encore vivantes qui ont bien connu Turing, il nous est possible d’avoir des aperçus concernant cette vie psychique. Ceux-ci sont très certainement parcellaires et peut-être même infidèles, mais ils sont néanmoins suffisants pour reconstituer les grandes lignes de ses pensées. Dès lors, pourquoi se priver d’une telle ressource quand c’est précisément du statut de la *psychologie* dans son rapport aux *mathématiques* qu’il est question ici ? Comment considérer qu’une science de la psychologie pourrait se constituer sans que soit prise en compte la dimension individuelle du psychisme ? Il nous a paru que l'occasion rare se présentait de montrer que le problème lié à la constitution d'une psychologie scientifique provenait précisément de la difficulté à articuler l'individuel et le général au sein d'une même théorie, *mais que cette articulation était possible*. Il nous restait alors la tâche difficile d'essayer de montrer comment une généralité pouvait se dégager de facteurs individuels si l'on faisait intervenir la constitution progressive des moyens linguistiques mis en place pour parvenir à exprimer ces facteurs. Il devenait dès lors possible d'interpréter ces facteurs d'un point de vue général et d'associer ainsi, par le biais de leur genèse, le particulier et le général.

Bibliographie

I- Ouvrages cités :

Ackermann W. (1928), "Zum Hilbertschen Aufbau der reellen Zahlen", Math. Annalen, vol. 99, p. 118-133; traduction anglaise dans [**van Heijenoort J. ed.** (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Cambridge], pp. 493-507.

Anderson A. R. ed. (1964), *Minds and Machines*, Prentice-Hall, Englewood Cliffs, N. J.; traduction française sous le titre *Pensée et Machine* (1983), Champ Vallon, Seyssel, pp. 39-67.

Andler D.,

- (1986), "Les sciences de la cognition" dans [**Hamburger J. dir.** (1986),*La philosophie de sciences aujourd'hui*, , Gauthier-Villars, Paris], pp. 131-167.
- (1991), "Representations : beyond the pro and con", manuscrit non-publié.
- (1992), "Calcul et représentations : les sources", dans [**Andler D. dir.** (1992), *Introduction aux sciences cognitives*, Folio Essais, Gallimard, Paris], pp. 19-20.
- (1992), **éd.**, *Introduction aux sciences cognitives*, Folio Essais, Gallimard, Paris.

Arbib M. (1972), *The Metaphorical Brain; An Introduction to Cybernetics as Artificial Intelligence and Brain Theory*, Wiley-Interscience, New York.

Babbage C. (1864), *Passages from the Life of a Philosopher*, London, Longman, reprint dans [**Morrison P.** et **Morrison E. eds.** (1961), *On the Principles and Developpment of the Calculator and other seminal writings*, by Charles Babbage and others, Dover Publications, New-York], pp. 3 à 157.

Barwise J.,

- (1977), **ed**., *Handbook of Mathematical Logic*, North-Holland, Amsterdam.
- (1980), **Barwise J.**, **Keisler J. J.** et **Kuchen K. eds.**,*The Kleene Symposium*, Amsterdam, North-Holland Publ. Co.

Benveniste E.,

- (1966), "Les niveaux de l'analyse linguistique" dans [**Benveniste E.** (1966), *Problèmes de linguistique générale*, I, Gallimard, Paris], pp. 119-131.
- (1974), "L'appareil formel de l'énonciation" dans [**Benveniste E.** (1974), *Problèmes de linguistique générale*, II, Gallimard, Paris], pp. 79-88.

- (1974), “La forme et le sens dans le langage” dans [**Benveniste E.** (1974), *Problèmes de linguistique générale*, II, Gallimard, Paris], pp. 215-238.

- (1974), “Sémiologie de la langue” dans [**Benveniste E.** (1974), *Problèmes de linguistique générale*, II, Gallimard, Paris, pp. 43-66.

Bion W. R. (1962), *Learning from Experience*, Basic Books, New York; traduction française sous le titre *Aux Sources de l’expérience*, P.U.F, Paris, 1979.

Block N.,

- (1980), **ed.**, *Readings in the Philosophy of Psychology*, I, Harvard University Press, 1980.

- (1980), “Troubles with Functionalism” dans [**Block N. ed.** (1980), *Readings in the Philosophy of Psychology*, I, Harvard University Press], pp. 268-305.

- (1980), “What is Functionalism ?” dans [**Block N. ed.** (1980), *Readings in the Philosophy of Psychology*, I, Harvard University Press], pp. 171-184.

Boden M. ed. (1990), *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford.

Borel E.,

- (1921), *Compte-Rendus de l’Académie des Sciences*, t. 173, p. 1304-1305, cité dans [**Saint-Sernin B.** (1973), *Les Mathématiques de la Décision*, Presses Universitaires de France, Paris], p. 176.

- (1934), Allocution à l’Académie des Sciences, séance du 17 décembre 1934, reproduit dans [**Fréchet M. éd.** (1967), *Borel, philosophe et homme d’action*, Gauthiers-Villars, Paris], pp. 353-365.

Bouveresse J. (1971), *La parole malheureuse*, Editions de Minuit, Paris.

Bowden ed. *Faster Than Thought*,

Brouwer L. E. J.,

- (1975), *Collected Works*, I, North-Holland, Amsterdam.

- (1907), “Over de Grondslagen der Wiskunde [On the Foundations of mathematics]”, *L. E. J. Brouwer Collected Works*, I, North-Holland, 1975, p. 83, cité dans [**Largeault J.** (1993), *Intuition et intuitionisme*, Vrin, Paris], p. 147.

- (1912), “Intuitionism and Formalism”, traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 39-53.

- (1948), “Consciousness, Philosophy and Mathematics”, traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 419-440.

- (1952), “Historical Background, Principles and Methods of Intuitionism” traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 445-463.

- (1953), “Points and Space”, *Collected Works*, I, p. 523, traduction française dans [**Largeault J.** (1993), *Intuition etntuitionisme*, Vrin, Paris], pp.191-220.

- (1981),*Cambridge Lectures on Intuitionism*, Cambridge University Press, Cambridge.

Bruter C. P. (1971), *Sur la nature des mathématiques*, Gauthier-Villars, Paris.

Butterworth ed. (1967), *Key Papers : Brain Physiology and Psychology*, University Park Press, Manchester, England.

Cantor G.,

- (1887), “Mitteilungen zur Lehre vom Transfiniten”, section VIII, pp. 411-412, cité dans [**Dauben J. W.** (1979), *Georg Cantor, His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, Mass.], p. 221.

- (1892), “Über eine elementare Frage der Mannigsfltigkeitslehre”, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 1, pp. 75-78, traduction française dans [**Rivenc F.** et **de Rouilhan P. dirs.** (1993), *Logique et Fondement des mathématiques*, Anthologie (1850-1914), Payot, Paris], pp. 200-203.

Carpenter B. E. et **Doran R. W. eds.** (1986), *A. M. Turing’s ACE report of 1946 and other papers*, MIT Press, Cambridge, Massachussetts.

Cassirer E. (1953), *Philosophie der symbolischen Formen*, Yale University Press; traduction française *La philosophie des formes symboliques*, I, Editions de Minuit, Paris, 1972.

Changeux J.-P. et **Connes A.** (1989), *Matière à pensée,* ed. O. Jacob, Paris.

Chevalley C. (1992), “Continu et Discontinu dans la construction de la théorie quantique. Un exemple.”, dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, , Springer-Verlag, Paris], pp. 327-337.

Church A.,

- (1937), Compte-rendu à l’article de Turing de 1936, publié dans le Journal of Symbolic Logic (2), pp. 42-43.

- (1938), “An unsolvable problem of Elementary Number Theory”, American Journal of Mathematics, 58, reprint dans [**Davis M. ed.** (1965), *The Undecidable*, Raven Press, Hewlett, New York], pp. 88-107,

Couturat L. (1901), *La logique de Leibniz,* Félix Alcan, Paris, réed. Olms Verlag, Hildesheim, 1985.

Dauben J. W. (1979), *Georg Cantor, His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, Mass.

Davis M.,

- (1965), **ed.**, *The Undecidable* Raven Press, Hewlett, New York.
- (1982), "Why Gödel didn't have Church's Thesis", Information and Control 54, pp. 3-24.

Dedekind R. (1888), *Continuité et nombres irrationnels*, trad. franç. par J. Milner revue par H. Sinaceur, Les cahiers d'Ornicar, Paris, 1978.

Delahaye J.-P. (1993), "Formulations mathématiques de la question : le monde est-il récursif ?", Cahiers du Crea n°15, Méthodologie de la science empirique (2) sous la direction d'Alain Boyer, Ecole Polytechnique, Paris, pp. 185-225.

Denis M.,

- (1993), "Pour les représentations", dans [**Denis M. et Sabah G. dirs.**, *Modèles et Concepts pour la science cognitive*, hommage à J. -F. Le Ny, Presses Universitaires de Grenoble, Grenoble, 1993], pp. 95-106.
- (1993), **Denis M. et Sabah G. dirs.**, *Modèles et Concepts pour la science cognitive*, hommage à J. -F. Le Ny, Presses Universitaires de Grenoble, Grenoble, 1993.

Dennett D. (1981), "Three Kinds of Intentional Psychology" dans [**Healey R. A. ed.**, *Reduction, Time and Reality : Studies in the Philosophy of natural sciences*, Cambridge, Cambridge University Press].

Dreyfus H. (1972), *What Computers Can't Do*, Harper and Row, New York; traduction française *Intelligence Artificielle: mythes et limites*, Flammarion, Paris, 1984.

Dubucs J.-P. (1991), "La philosophie de Kurt Gödel", L'âge de la science, Gallimard, Paris.

Evans C. R. et **Robertson A. D. G. eds.** (1968), *Key Papers : Cybernetics*, University Park Press, Manchester, England.

Feferman S. (1988), "Turing in the Land of O(z)", dans [**Herken R. ed.** (1988), *The Universal Turing Machine*, Oxford University Press], pp. 113-147.

Feigenbaum et **Feldman eds.** (1963), *Computers and Thought,* McGraw-Hill, New-York.

Fenstad J. E. ed. (1970), *Collected Works*, ed. **Fenstad J. E.**, Universitetsforlaget, Oslo.

Fodor J. A.,

- (1975), *The Language of Thought*, Harvard University Press, Harvard.
- (1983), *The Modularity of Mind*, MIT Press, traduction française sous le titre *La modularité de l'esprit*, Minuit, Paris, 1986.

Frege G.,

- (1879), *Begriffschrift*, traduction française dans [**F. Rivenc et P. de Rouilhan dirs.** (1992), *Logique et Fondements des mathématiques*, Anthologie (1850-1914), Payot, Paris], pp. 93-129.
- (1918-1919), "Recherches logiques. 1. la Pensée", traduction française dans [**Frege G.** (1971), *Ecrits logiques et philosophiques*, Le Seuil, Paris], pp. 170-195.
- (1971), *Ecrits logiques et philosophiques*, Le Seuil, Paris.

French R. M. (1990), "Subcognition and the limits of the Turing test", Mind, 99, pp. 53-65.

Ganascia J.-G (1990), *L'âme-machine; les enjeux de l'intelligence artificielle*, Le Seuil, Paris.

Gandy R. O.,

- (1980), "Church's thesis and principles for mechanisms" dans [**Barwise J.**, **Keisler J. J.** et **Kuchen K. eds.** (1980),*The Kleene Symposium*, , Amsterdam, North-Holland Publ. Co.], pp. 123-145.
- (1988), "The Confluence of Ideas in 1936" dans [**Herken R. ed.** (1988), *The Universal Turing machine*, Oxford University Press], pp. 55-110.

Gardies J.-L. (1988), *L'héritage épistémologique d'Eudoxe de Cnide*, *un essai de reconstitution*, Vrin, Paris.

Gardner H. (1985), *The Mind's New Science*, Basic Books, New-York.

Gödel K.,

- (1986), *Collected Works*, t. I, Oxford University Press, Oxford.
- (1990), *Collected Works*, t. II, Oxford University Press, Oxford.
- (1949), "An example of a new type of cosmological solutions of Einstein's field equations of gravitation", Review of modern physics, 21, republié dans [**Gödel K.** (1990), *Collected Works*, t. II, Oxford University Press, Oxford], pp. 190-198.
- (1952), "What is Cantor's continuum problem ?", Bull. Am. Math. Soc., vol. 58, 1952 et dans [**Benacerraf** et **Putnam eds.** (1983), *Philosophy of mathematics*, CUP, Cambridge, 2nd edition, 1983] pp. 470-485, traduction

française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 509-531.

- (1972a), “Some Remarks on the Undecidability Results”, reproduit dans [**Gödel K.** (1990),*Collected Works*, t. II, Oxford University Press, Oxford], pp. 305-306.

Goldstine H. (1972), *The Computer from Pascal to Von Neumann*, Princeton University Press, Princeton.

Gonseth F. (1964), *Le problème du temps*, Editions du Griffon, Neuchatel.

Good I. J. (1979), “Introductory Remarks for the article in Biometrika 66, “A. M. Turing’s Statistical Work in World War II” publié dans [**Turing A. M.** (1992), *Collected Works of A. M. Turing*, vol. 1 “Pure Mathematics”, North-Holland], pp. 215-217.

Goodman N. (1977), *The Structure of Appearance*, Dordrecht-Boston.

Goody J. (1977), *La raison graphique; la domestication de la pensée sauvage*, Éditions de Minuit, Paris, 1979.

Graubard S. T. ed. (1988), *The Artificial Intelligence Debate, False Starts, Real Foundations*, MIT Press, Cambridge, Mass..

Hacking I.,

- (1975), *The Emergence of Probability*, Cambridge University Press, Cambridge.
- (1990), *The Taming of Chance*, Cambridge University Press, Cambridge.

Hadamard J. (1945), *Essai sur la psychologie de l’invention dans le domaine mathématique*, traduction française, Gauthiers-Villars, Paris, 1959.

Hallet M. (1984), *Cantorian set theory and limitation of size*, Clarendon Press, Oxford.

Hamburger J. dir. (1986), *La philosophie de sciences aujourd’hui*, Gauthier-Villars, Paris.

Hao W.,

- (1965), “Games, Logic and Computers”, Scientific American, 213 (5), pp. 98-106.
- (1974), *From mathematics to philosophy*, Routledge and Kegan Paul, London.
- (1981a), “Some Facts about Kurt Gödel, Journal of Symbolic Logic, XLVI, pp. 653-659.
- (1981b) *Popular Lectures in Mathematical Logic*, Van Nostrand Reinhold Company, New York.

- (1987), *Reflections on Kurt Gödel*, MIT Press, Cambridge, Mass.

Haugeland J.,

- (1981), "Semantic Engines : an Introduction to Mind Design" dans [**Haugeland J. ed.** (1981), *Mind Design*, MIT Press, Mass. 1981], pp. 1-34.
- (1981), éd., *Mind Design*, MIT Press, Mass.
- (1985), *Artificial Intelligence, the very idea,* traduction française *L'esprit dans la machine, Fondements de l'intelligence artificielle*, ed. Odile Jacob, Paris, 1989.

Healey R. A. (1981), *Reduction, Time and Reality : Studies in the Philosophy of natural sciences*, Cambridge, Cambridge University Press.

Heims S. J. (1980), *John von Neumann and Norbert Wiener; From Mathematics to the Technologies of Life and Death*, MIT Press, Cambridge, Massachussets.

Herken R. ed. (1988), *The Universal Turing Machine*, Oxford University Press, Oxford.

Hilbert D.,

- (1922), "Neubegründung der Mathematik. Erste Mitteilung", dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], p. 107-130.
- (1925), "Über das Unendliche", traduction française dans [**Largeault J. éd.** (1972), *Logique mathématique,* textes, Armand Colin, Paris], pp. 215-245.
- (1927), "Die Grundlagen der Mathematik", traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 131-144.
- (1928), "Probleme der Grundlegung der Mathematik", traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 175-185.

Hinsley F. H. (1979-1991), *British Intelligence in the Second World War*, vol I -VI, Her Majesty's Stationery Office Publications, London.

Hobbes, *Leviathan*, (1651), Dent, London, 1973.

Hodges A.,

- (1983), *Alan Turing, The Enigma of Intelligence*, Unwin Paperbacks, London.
- (1988), "Alan Turing and The Turing Machine" dans [**Herken R. ed.** (1988),*The Universal Turing machine*, Oxford University Press], pp. 3-15.

Hofstadter D.,

- (1979), *Escher et Bach : an Eternal Golden Braid* , Basic Books, New York, 1979, traduction française*Gödel, Escher et Bach*, Interéditions, Paris, 1985.

- (1981), **Hofstadter D. et Dennett D. C. eds.**, *The Minds'I*, Basic Books, New York; traduction française sous le titre*Vues de l'esprit*, InterEditions, Paris, 1987.

Hulbert A. et **Poggio T.** (1988), "Making machines (and artificial intelligence) see" dans [**Graubard S. T. ed.** (1988), *The Artificial Intelligence Debate, False Starts, Real Foundations,* Cambridge, Mass : MIT Press], pp. 213-239.

Jastrow R. (1981), *Au-delà du cerveau*, traduction française, Hachette, 1984.

Kleene S. C.,

- (1952), *Introduction to Metamathematics*, North-Holland,, Amsterdam.
- (1988), "Turing's Analysis of Computability and Major Applications of It", dans [**Herken R. ed.** (1988), *The Universal Turing Machine*, Oxford University Press, 1988], pp. 17-54.

Kneale W. et **Kneale M.**, *The Development of Logic*, Clarendon Press, Oxford, 1962.

Köhler W. (1964), *Psychologie de la forme*, Gallimard, Paris.

Kreisel G.,

- (1958), "Hilbert's programme", traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 485-486.
- (1970), "Church's Thesis : a Kind of Reducibility Axiom for Constructive Mathematics" dans [**Myhill, Kino, Vesley eds.** (1970), *Intuitionism and Proof Theory*, North-Holland] pp. 121-150.
- (1987), "Church's Thesis and the Ideal of Informal Rigour", Notre Dame Journal of Symbolic Logic, 28, 4, pp. 499-519.
- (1988), "Review of Gödel's 'Collected Works, Volume I'", Notre Dame Journal of Formal Logic, 29, 1, pp. 134-171.

Largeault J.,

- (1972), **éd.**, *Logique mathématique,* textes, Armand Colin, Paris.
- (1985), *Principes de philosophie réaliste*, Klincksiek, Paris.
- (1988), *Principes classiques d'interprétation de la nature*, Publications de l'Institut Interdisciplinaire d'Etudes Epistémologiques, Vrin, Paris.
- (1992), **éd.**, *Intuitionisme et théorie de la démonstration*, Vrin, Paris.
- (1992), *L'intuitionisme*, Coll. Que Sais-je ? n° 2684, Presses Universitaires de France, Paris.
- (1993), *Intuition et Intuitionisme*, Vrin, Paris.
- (1993), *La logique*, Coll. Que Sais-je ? n° 225, Presses Universitaires de France, Paris.

Lautman A., “Considérations sur la logique mathématique” dans [**Lautman A.** (1977), *Essai sur l’unité des mathématiques et divers écrits*, Union Générale d’Édition, Paris, 1977], pp. 305-315.

Leiber J. (1991), *An invitation to Cognitive Science*, Basil Blackwell, Oxford.

Leibniz,

- (1680), “Discours touchant la méthode de la certitude et l’art d’inventer”, *die Philosophischen Schriften*, **Gerhardt ed.** (1890), Olms Verlag, Hildesheim, Band VII, pp. 174-183.
- (1678), “Inventorium Mathematicum”, *Mathematische Schriften*, **Gerhardt ed.** (1863), Olms Verlag, Hildesheim, Band VII, pp. 13-17.
- (1710), *Nouveaux Essais*, *die Philosophischen Schriften*, **Gerhardt ed.** (1890), Olms Verlag, Hildesheim, Band V, pp. 39-509.

Lucas J. R. (1961), “Minds, Machines and Gödel”, Philosophy, XXXVI; traduction française dans [**Anderson A. R. ed.** (1964), *Pensée et Machine*, Champ Vallon, Seyssel, 1983], pp. 81-97.

Maddy P. (1990), *Realism in Mathematics*, Clarendon Press, Oxford.

Marr D. (1990), “Artificial Intelligence : a personal view” dans [**Partridge D.** et **Wilks Y. eds.** (1990), *The Foundations of Artificial Intelligence*; *a sourcebook*, Cambridge University Press, Cambridge], pp. 97-107.

Martin R. (1964), *Logique contemporaine et formalisation*, Presses Universitaires de France, Paris.

Merleau-Ponty J. (1965), *Cosmologie du XXème siècle*, Gallimard, Paris.

Messiah A. (1969), *Traité de mécanique quantique,* 2ème édition, Dunod, Paris.

Meyering T. C. (1989), *Historical roots of cognitive science*, *the rise of a cognitive theory of perception from Antiquity to the 19th century*, Kluner Academic Publishers, Dordrecht.

Meyerson E. (1951), *Identité et réalité*, Vrin, Paris.

Michie D.,

- (1974), *On Machine Intelligence*, a Halsted Press Book, John Wiley and Sons, New York.
- (1993), “Turing’s Test and conscious thought”, Artificial Intelligence, 60, pp. 1-22.

Minsky M. L. (1967), *Computation : Finite and Infinite machines*, Prentice-Hall, Inc. Englewood Cliffs, N. J..

Morrison P. et **Morrison E. eds.** (1961), *On the Principles and Developpment of* *the Calculator and other seminal writings*, by Charles Babbage and others, Dover Publications, New-York.

Mosconi J. (1989), *La constitution de la théorie des automates,* thèse de doctorat d'Etat, Université de Paris I.

Myhill, Kino, Vesley eds. (1970), *Intuitionism and Proof Theory*, North-Holland, Amsterdam.

Nelson R. J. (1987), "Church Thesis and Cognitive Science", Notre-Dame Journal of Symbolic Logic, XXVIII-4.

Ockam, *Somme de Logique*, T. E. R., Mauvezin, 1988.

Panza M. (1992), "De la continuité comme concept au continu comme objet", dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, Springer-Verlag, Paris], pp. 16-30.

Partridge D. et **Wilks Y. eds.** (1992), *The Foundations of Artificial Intelligence*; *a sourcebook*, Cambridge University Press, Cambridge.

Pascal, *Œuvres Complètes*, Le Seuil, Paris, 1963.

Penrose R.,

- (1988), "On Physics and Mathematics of Thought" dans [**Herken R. ed.** (1988), *The Universal Turing machine*, Oxford University Press], pp. 491-522.
- (1989), *The Emperor's New Mind*, Oxford University Press, Oxford.

Petitot J. (1985), *Morphogenèse du Sens*, I, Presses Universitaires de France, Paris.

Platon, *Théétète*, Les Belles Lettres, Paris, 1926.

Poincaré H.,

- (1902), *La Science et l'hypothèse*, Flammarion, Paris.
- (1905), *La valeur de la science,* Flammarion, Paris.

Post E. (1936), "Finite Combinatory Processes - Formulation I", Journal of Symbolic Logic, I, réédité dans [**Davis M. ed.** (1965), *The Undecidable*, Raven Press, Hewlett, New York], pp. 288-291.

Poundstone W. (1987), *The Recursive Universe*, Oxford University Press, Oxford.

Pour-El M.-B. et **Richards I.** (1979), "A computable differential equation which posesses no computable solution", Ann. Math. Log, 17, pp. 61-90.

Pylyshyn Z. (1984), *Computation and Cognition, Toward a Foundation for Cognitive Science,* A Bradford Book, MIT Press, Cambridge, Massachusetts.

Ramunni J. (1989), *La physique du calcul; Histoire de l'ordinateur*, Hachette, Paris.

Rastier F. (1991), *Sémantique et recherches cognitives*, Presses Universitaires de France, Paris.

Rejewski M. (1980), “Jak matematycy polscy rozszyfrowali Enigme”, Annales de la Société Mathématique Polonaise, deuxième série, *Wiadomosci Matematyczne*, Volume 23, pp. 1-28; traduction anglaise “How Polish Mathematicians Deciphered the Enigma”, Annals of the History of Computing, Volume 3, Number 3, July 1981, pp. 213-234.

Richard J. (1905), “Les principes des mathématiques et le problème des ensembles”, Revue générale des sciences pures et appliquées, 16, 541, reprint dans [**Rivenc F.** et **de Rouilhan P. dirs.** (1992), *Logique et Fondement des mathématiques*, Anthologie (1850-1914), , Payot, Paris], pp. 271-275.

Rivenc F. et **de Rouilhan P. dirs.** (1992), *Logique et Fondement des mathématiques*, Anthologie (1850-1914), Payot, Paris.

Rosen R. (1988), “Processes and Natural Laws”, dans [**Herken R. ed**. (1988),*The Universal Turing machine*, Oxford University Press, Oxford], pp. 523-537.

Ruelle D. (1991), *Ordre et Chaos*, ed. O. Jacob, Paris.

Russell B.(1902), Lettre à Frege, traduction française, dans [**Rivenc F.** et **de Rouilhan P. dirs.** (1992), *Logique et Fondements des mathématiques*, Anthologie (1850-1914), Payot, Paris], pp. 240-241.

Ruyer B. (1990), *Logique*, P.U.F, Paris.

Ruyer R. (1966), *Paradoxes de la conscience et limites de l’automatisme*, Albin Michel, Paris.

Saint-Sernin B. (1973), *Les Mathématiques de la Décision*, Presses Universitaires de France, Paris.

Salanskis J.-M.,

- (1991), *L’herméneutique formelle*, CNRS, Paris.
- (1992), “Le destin du modèle de Cantor-Dedekind” dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, Springer-Verlag, Paris], pp. 190-212.
- (1992), **Salanskis J.-M.** et **Sinaceur H. dirs.**, *Le Labyrinthe du Continu*, Springer-Verlag, Paris.

Simondon G., (1958), *Du mode d'existence des objets techniques,* Aubier, Paris.

Sinaceur H.,

- (1992), **Salanskis J.-M.** et **Sinaceur H. dirs.***Le Labyrinthe du Continu*, Springer-Verlag, Paris.
- (1994), *Jean Cavaillès, philosophie mathématique*, Presses Universitaires de

France, Paris.

Skolem T. (1970), “Sur la portée du Théorème de Löwenheim-Skolem” dans [**Fenstad J. E. ed.** (1970), *Collected Works*, Universitetsforlaget, Oslo], pp. 455-482.

Smullyan R. M. (1961), *Theory of Formal Systems*, Annals of Mathematics Studies n° 47, Princeton University Press, Princeton.

Thom R.,

- (1980), *Modèles mathématiques de la morphogenèse*, 2ème édition, Bourgois, Paris.
- (1983), *Paraboles et catastrophes*, Flammarion, Paris.

Turing A. M.,

- (1992) *Collected Works of A. M. Turing*, vol. 1 “Pure Mathematics”; vol. 4 “Morphogenesis”, North-Holland, Amsterdam.
- (1936), ‘On Computable Numbers with an Application to the Entscheidungproblem’, Proceedings of the London Mathematical Society, 42 : 230-265.
- (1939), “Systems of Logic based on Ordinals”, Proceedings of the London Mathematical Society, ser. 2, vol. 45 (1939), pp. 161-228, réédité dans [**Davis M. ed.**, (1965), *The Undecidable*, Raven Press, Hewlett, New York], pp. 154-222.
- (1947), “Lecture to the London Mathematical Society on 20 February 1947”, reprint dans [**Carpenter B. E.** et **Doran R. W. eds.** (1986), *A. M. Turing’s ACE report of 1946 and other papers*, MIT Press, Cambridge, Massachussetts], pp. 106-124.
- (1948), “Proposal for Development in the Mathematics Division of an Automatic Computing Engine (ACE)”, H. M. S. O., republié dans [**Carpenter B. E.** et **Doran R.W. eds.** (1986), *A. M. Turing ‘s ACE Report of 1946 and other Papers*, MIT Press, Cambridge], pp. 20-105.
- (1948), “Intelligent Machinery”, Executive Committee NPL, 1-20, H. M. S. O.; republié dans [**Evans C. R.** et **Robertson A. D. G. eds.** (1968), *Key Papers : Cybernetics*, University Park Press, Manchester, England], pp. 27-52.
- (1950), “Computing Machinery and Intelligence”, Mind, Vol. LIX. n° 236, pp. 433-460.
- (1952), “The Chemical Basis of Morphogenesis”, Phil. Trans. Roy. Soc. B 237, p. 24; republié dans [**Turing A. M.**, *Collected Works of A. M. Turing*, vol. 4, “Morphogenesis”, North-Holland, Amsterdam, 1992], pp. 1-72.

- (1953), “Digital Computers Applied to Games”, dans [**Bowden ed.** (1953), *Faster Than Thought*, Pitman, London].

- (1954), “Solvable and Unsolvable Problems”, Science News, 31, Penguin Science, pp. 1- 23.

Van Heijenoort J.,

- (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Cambridge.

- (1967), “Logic as calculus and Logic as language” dans [**Van Heijenoort J.** (1985), *Selected Articles*, Bibliopolis, Napoli], pp. 11-16.

- (1985), *Selected Articles*, Bibliopolis, Napoli, 1985.

Von Neumann J. (1966), *Theory of Self-Reproducing Automata*, ed. A. Burks, University of Illinois Press, Urbana and London.

Wagner P. (1994), *Machine et pensée : l’importance philosophique de l’informatique et de l’intelligence artificielle*, thèse de doctorat de philosophie de l’Université Paris-I.

Webb J. C.,

- (1980), *Mechanism, Mentalism and Metamathematics : an Essay on Finitism*, Reidel, Dordrecht.

- (1990), “Introductory note to 1972a” dans [**Gödel K.** (1990), *Collected Works*, vol. II, Oxford University Press, Oxford], pp. 281-304.

Weedon D. ed. (1992), *Systematic Pathology*, Gen. Ed. W. St. C. Symmens, 3rd edition, Volume 9., Churchill Livingstone, London.

Weyl H. (1921), “Sur la crise contemporaine des fondements des mathématiques”, traduction française dans [**Largeault J. éd.** (1992), *Intuitionisme et théorie de la démonstration*, Vrin, Paris], pp. 76-77.

Widlöcher D. (1993), “Croire en l’inconscient”, Nouvelle revue de psychanalyse, n°48, Gallimard, Paris, pp. 93-113.

Wiener N. (1964), *God and Golem, Inc.*, MIT Press, Cambridge, Massachusetts.

Yourgrau P. (1991), *The Disappearance of Time, Kurt Gödel and the Idealistic Tradition in Philosophy*, Cambridge University Press, Cambridge.

Zurek W. H. (1990), *Complexity, Entropy and the Physics*, Addison-Wesley Publishing Company.

II- Ouvrages consultés :

1. Philosophie des mathématiques

Barreau H. et **Hartong J. dirs.** (1989), *La mathématique non standard*, Presses du CNRS, Paris.

Benacerraf P. et **Putnam H. eds.** (1983), *Philosophy of mathematics*, Cambridge University Press, 2nd edition, Cambrige.

Borel E.,

- (1900), “A propos de l’”infini nouveau””, Revue philosophique, t. 48, pp. 383-390, réédité dans [**Fréchet M. éd.** (1967), *Borel, Philosophe et homme d’action*, Gauthiers-Villars, Paris], pp. 139-154.
- (1904), *Les paradoxes de l’infini*, Gallimard, Paris, 1904.
- (1907), “L’évolution de l’intelligence géométrique”, Revue de Métaphysique et de Morale, T. 15, n°6, pp. 273-283, réédité dans [**Fréchet M. éd.**, *Borel, Philosophe et homme d’action*, Gauthiers-Villars, Paris, 1967], pp. 24-31.
- (1908), Le continu mathématique et le continu physique”, Scientia, t. 6, pp. 21-35, réédité dans [**Fréchet M. éd.** (1967), *Borel, Philosophe et homme d’action*, Gauthiers-Villars, Paris, 1967], pp. 186-200.
- (1909), “La théorie des ensembles et les progrès récents de la théorie des fonctions”, Revue générale des sciences, t. 20, pp. 315-324, réédité dans [**Fréchet M. éd.** (1967), *Borel, Philosophe et homme d’action*, Gauthiers-Villars, Paris, 1967], pp. 155-175.
- (1914), L’infini mathématique et la réalité”, La revue du mois, t. 18, pp. 71-84, réédité dans [**Fréchet M. éd.** (1967), *Borel, Philosophe et homme d’action*, Gauthiers-Villars, Paris, 1967], pp. 176-185.
- (1936), La psychologie de l’invention dans le domaine de la science”, Organon International Review published by the Mianowski Institute for the Promotion of Science and Letters, vol. I, pp. 33-42, réédité dans [**Fréchet M. éd.** (1967), *Borel, Philosophe et homme d’action*, Gauthiers-Villars, Paris], pp. 323-331.
- (1952), *Les Nombres inaccessibles*, Gauthiers-Villars, Paris.

Brunschvicg L. (1930), *Les étapes de la philosophie mathématique*, Felix Alcan, Paris.

Bruter C. P. (1987), *De l'intuition à la controverse*, Paris, Blanchard.

Burali-Forti C. (1897), "Una questione sui numeri transfiniti" *Rendiconti del Circolo matematico di Palermo*, 11, pp. 154-164; traduction anglaise dans [**van Heijenoort J. ed.** (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Harvard], pp. 104-111.

Cantor G.,

- (1895/1897), "Beiträge zur Begründung der Transfiniten Mengenlehre", Mathematische Annalen; traduction française dans [**Marotte F.** (1899), Mémoires de la Société des Sciences physiques et naturelles de Bordeaux, tome III, (5è série), "Sur les fondements de la théorie des ensembles transfinis", réédition par Editions Jacques Gabay, Paris, 1989].
- (1873/1899), **Cantor** et **Dedekind**, *Correspondance*, traduction française dans [**Cavaillès éd.** (1962), *Philosophie mathématique*, Hermann, Paris], pp. 179-250.

Cavaillès J.,

- (1937), *Méthode axiomatique et formalisme*, Hermann, Paris, réed. 1981.
- (1947), *Sur la logique et la théorie de la science*, Vrin, Paris.
- (1962), *Philosophie mathématique*, Hermann, Paris.

Caveing M.,

- (1982a), "Quelques remarques sur le traitement du continu dans les *Eléments* d'Euclide et la *Physique* d'Aristote", dans [**Guénard F.** et **Lelièvre G. dirs.** (1982), *Penser les mathématiques*, Le Seuil, Paris], pp. 145-166.
- (1982), *Zénon d'Elée : prolégomènes aux doctrines du continu*, Vrin, Paris.

Chaitin G. J.,

- (1975), "Randomness and mathematical proof", Scientific American, n° 232, pp. 47-52.
- (1979), "Toward a mathematical definition of "Life"" dans [**Levine R.** et **Tribus M. eds.** (1979) *The Maximum Entropy Formalism*, MIT Press, Cambridge, Massachussetts], pp. 477-498.
- (1988), "Le hasard en théorie des nombres", Pour la Science, n° 131, Septembre 1988, pp. 82-87.

Chirollet J.-C. (1989), “Le continu mathématique du troisième ordre chez Henri Poincaré”, dans [**Barreau H.** et **Hartong J. dirs.** (1989), *La mathématique non-standard*, presses du CNRS, Paris], pp. 83-116.

Courant R. et **Robins H.** (1941), *What is mathematics ?*, Oxford University Press, Oxford.

Cournot A.A. (1847), *De l’origine et des limites de la correspondance entre l’algèbre et le géométrie*, réédition dans [**Cournot A.A.** (1989), *Œuvres Complètes*, Tome VI/2, J. Vrin, Paris].

Couturat L.,

- (1896), *De l’infini mathématique*, Felix Alcan, Paris; réed Blanchard, Paris, 1973.
- (1900), “Le continu mathématique”, <u>Revue de métaphysique et de morale</u>.

Dahan-Dalmedico A. et **Peiffer J.** (1982), *Une histoire des mathématiques,* ed. Etudes Vivantes, Paris; réed. Le Seuil, Paris, 1986.

Dantzig T. (1930), *Number, the Language of Science*, Macmillan, New-York.

Delahaye J.-P. (1991), “Le réalisme en mathématiques et en physique”, <u>Pour la Science</u>, n° 159, pp. 34-42.

Desanti J.-T. (1967), "Une crise de développement exemplaire : la “découverte” des nombres irrationnels", dans [**Piaget J. dir.** (1967), *Logique et connaissance scientifique*, Encyclopédie de la Pléiade, Gallimard, Paris], pp. 439-464.

Detlefsen M. (1966), *Hilbert’s program : an essay on mathematical instrumentalism*, Reidel Publishing Company, Dordrecht.

Ekeland I. (1984), *Le calcul, l’imprévu : les figures du temps de Kepler à René Thom*, Le Seuil, Paris.

Fang J. (1976), *The Illusory Infinite -a theology of mathematics*, Paideia Press.

Feferman S.,

- (1987a), “Infinity in mathematics : Is Cantor necessary ?” dans [*Infinity in Science*, Instituto della Encyclopedia italiana, Roma], pp. 151-209.
- (1987b), “Weyl vindicated : “Das Kontinuum” 70 years later”, dans [Atti del congresso : Temi e prospettiva della logica e della filosofia della scienza, 7-10 Gennaio 1987, vol 1, CLUE B, Bologna], pp. 59-93.

Fraenkel A. A., **Bar-Hillel Y.** et **Levy A.** (1973), *Foundations of set theory*, 2nd edition, North-Holland, Amsterdam..

Fréchet M.,

- (1942), "L'arithmétique de l'infini", Science et Vie, vol. LXVII, pp. 38-47, réédité dans [**Fréchet M.** (1955), *Les mathématiques et le concret*, Presses Universitaires de France, Paris, pp. 401-413].
- (1967), **éd.,** *Borel, Philosophe et homme d'action*, Gauthiers-Villars, Paris.

Gardies J.-L.,

- (1984), *Pascal entre Eudoxe et Cantor*, Vrin, Paris, 1984.
- (1988), *L'héritage épistémologique d'Eudoxe de Cnide*, un essai de reconstitution, Vrin, Paris.

Gonseth F.,

- (1936), *Les mathématiques et la réalité : essai sur la méthode axiomatique*, Blanchard, Paris.
- (1945), *La géométrie et le problème de l'espace*, éditions du Griffon, Neuchatel.

Guénard F. et **Lelièvre G. dirs.** (1982), *Penser les mathématiques*, Le Seuil, Paris.

Hadamard J. (1905), "Cinq lettres sur la théorie des ensembles" (Baire, Borel, Hadamard, Lebesgue) dans [**Hadamard J.** (1968), *Œuvres*, I, Editions du CNRS, Paris], pp. 335-347.

Harré R. ed. (1969), *Scientific Thought 1900-1960, a selective survey*, Clarendon Press, Oxford.

Hartong J.,

- (1984), **Hartong J.** et **Reeb G.,** "Intuitionnisme 84", dans [**Barreau** et **Hartong dirs.** (1989), *La mathématique non-standard,* , presses du CNRS, Paris], pp. 213-273.
- (1987), "Le Continu et l'Ordinateur", L'Ouvert, 46, pp. 13-27.
- (1989), **Barreau H.** et **Hartong J. dirs.**, *La mathématique non standard*, Presses du CNRS, Paris.

Heinzmann G.,

- (1985), *Entre intuition et analyse : Poincaré et le concept de prédicativité,* Blanchard, Paris.

- (1986), *Russell, Poincaré, Zermelo et Peano*, Blanchard, Paris.

Kreisel G. (1969), “Foundations of mathematics : 1900-1950” dans [**Harré R. ed.** (1969), *Scientific Thought 1900-1960, a selective survey*, Clarendon Press, Oxford], pp. 4-13.

Largeault J. (1990), “Formalisme, intuitionnisme en philosophie des mathématiques”, Revue philosophique, n°3, Presses Universitaires de France, Paris, Juillet-Septembre], pp. 521-546

Lautman A. (1977), *Essai sur l’unité des mathématiques et divers écrits*, 10/18, Union Générale d’Edition, Paris.

Le Roy E. (1960), *La pensée mathématique pure*, Presses Universitaires de France, Paris.

Levine R. et **Tribus M. eds.** (1979), *The Maximum Entropy Formalism*, MIT Press, Cambridge, Massachussetts.

Lobry C.(1989) , *Et pourtant... ils ne remplissent pas N !*, Aleas Editeur, Lyon.

Maddy P.,

- (1988a), “Believing the axioms I”, The Journal of Symbolic Logic, vol 53, n°2.

- (1988b), “Believing the axioms II”, The Journal of Symbolic Logic, volume 53, n°3.

- (1989), “The roots of contemporary platonism”, The Journal of Symbolic Logic, vol 54, n°4.

Petitot J. (1992), “Continu et Objectivité; La bimodalité objective du continu et le platonisme transcendantal” dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, Actes du Colloque de Cerisy de 1990, Springer Verlag, Paris], pp. 237-263.

Piaget J. (1967), “Les problèmes principaux de l’épistémologie des mathématiques”, dans [*Logique et connaissance scientifique*, **Piaget J. dir.** (1967), Gallimard, Paris], pp. 554-596.

Salanskis J.-M.,

- (1985), “Continu et Discret” dans [Encyclopedia Universalis.

- (1989), “Le potentiel et le virtuel”, dans [**Barreau H.** et **Hartong J. dirs.** (1989), *La mathématique non standard*, Presses du CNRS, Paris], pp. 275-303.

- (1992), “Le destin du modèle de Cantor-Dedekind” dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, Springer Verlag, Paris], pp. 190-212.

- (1992), **Salanskis J.-M.** et **Sinaceur H. dirs.**, *Le Labyrinthe du Continu*, Springer Verlag, Paris, 1992.

Ulam S. (1958), “John Von Neumann”, American Mathematical Society Bulletin, 64, 3(P2), pp. 1-49.

van Heijenoort J. (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Harvard.

Wallet G. (1992), “Réflexions sur l’Objectivité en mathématiques” dans [**Salanskis J.-M.** et **Sinaceur H. dirs.** (1992), *Le Labyrinthe du Continu*, Springer Verlag, Paris], pp. 230-238.

Weyl H. (1949), *Philosophy of mathematics and natural science*, Princeton University Press, Princeton.

Zermelo E.,

- (1904), “Beweis, dass jede Menge wohlgeordnet werden kann”, Mathematische Annalen, 1904, pp. 514-516; traduction anglaise dans [**van Heijenoort J. ed.** (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Cambridge, Massachussetts, pp. 139-141].

- (1908), “Neuer Beweis für die Möglichkeit einer Wohlordnung”, Mathematische Annalen, 1908, pp. 107-128; traduction française dans [**De Rouilhan P.** et **Rivenc F. dirs.** (1992), *Logique et Fondements des mathématiques*, (1850-1914), Payot, Paris, pp. 335-366].

- (1908), “Untersuchugen über die Grundlagen der Megenlehre” I, Mathematische Annalen, 1908; traduction française dans [**De Rouilhan P.** et **Rivenc F. dirs.** (1992), *Logique et Fondements des mathématiques*, (1850-1914), Payot, Paris, 1992, pp. 367-378].

2. Philosophie de la logique

Andler D. (1985), “Logique mathématique” dans [Encyclopedia Universalis].

Barwise J., **Keisler J.J** et **Kuchen K. eds.** (1980), *The Kleene Symposium*, North-Holland Publ. Co., Amsterdam..

Barwise J. et **Etchemendy J.** (1989), “Model-Theoretic Semantics”, chapitre 6 de [**Posner ed.** (1989), *Foundations of Cognitive Science*, MIT Press, Cambridge, Massachussetts], pp. 207-263.

Bouleau N., **Girard J.-Y.** et **Louveau A.** (1983), “Cinq conférences sur l’indécidabilité”, Presses de l’Ecole Nationale des Ponts et Chaussées, Paris.

Couturat L.,

- (1901), *La logique de Leibniz,* Felix Alcan, Paris, 1901, rééd. Olms Verlag, Hildesheim, 1985.
- (1906), *Les principes des mathématiques*, rééd. Blanchard, Paris, 1980.

Delahaye J.-P.,

- (1988), “Une extension spectaculaire du théorème de Gödel : l’équation de Chaitin”, La Recherche, n° 200, Juin 1988, pp. 860-862.
- (1991), “Le concept de suite aléatoire et la thèse de Church”, Séminaire de Philosophie et mathématiques de l’E. N. S., édité par l’I.R.E.M. Paris-Nord, Université Paris-XIII, n°76.

Gödel K. (1944), “Russell’s mathematical logic” dans [**Benacerraf** et **Putnam eds.** (1983), *Philosophy of mathematics,* CUP, Cambridge], pp. 211-232.

Haack S. (1978), *Philosophy of logics*, Cambridge University Press, Cambridge.

Hacking I. (1979), "What is logic ?", The Journal of Philosophy, volume LXXVI, n° 6, pp. 285-319.

De Rouilhan P. et **Rivenc F. dirs.** (1992)*, Logique et Fondements des mathématiques*, Anthologie (1850-1914), Payot, Paris.

Ladrière J.,

- (1957), *Les limitations internes des formalismes*, Gauthier-Villars, Paris 1957.

- (1967), “Les limites de la formalisation”, dans [**Piaget, J. dir.** (1967), *Logique et connaissance scientifique*, Gallimard, Paris, pp. 312-333].

- (1968), “Le théorème de Löwenheim-Skolem”, Cahiers pour l’analyse, Presses de l’Ecole Normale Supérieure, Paris, pp. 108-130.

Nagel E. et **Newman J. R.** (1958), *Gödel’s proof*, New York University Press, New York.

Posner ed. (1989), *Foundations of Cognitive Science*, MIT Press, Cambridge, Massachussetts.

Russell B. (1903), *Principles of mathematics,* Allen and Unwin, London.

Sinaceur H., (1988), “*Ars inveniendi* et théorie des modèles”, Dialogue, XXVII.

Stewart I. (1988), “The ultimate undecidability” , Nature, vol 332, pp. 115-116.

Van Heijenoort J. ed. (1967), *From Frege to Gödel : a source book in mathematical logic,* Harvard University Press, Cambridge, Massachussetts.

Weyl H. (1929), “Consistency in Mathematics”, The Rice Institute Pamphlet, vol. XVI, n° 4, pp. 245-265.

3. Machines de Turing, théorie de la calculabilité

Andler D. (1984), “La Machine Universelle”, Science et Avenir, n° 49 hors série, pp. 43-48.

Arbib M. A. (1988), “From Universal Turing Machines to Self-Reproduction”, dans [**Herken R. ed.** (1988),*The Universal Turing machine*, Oxford University Press, pp. 177-189].

Caianiello E. R. ed. (1966), *Automata Theory*, Academic Press, London.

Davis M.,

- (1958), *Computability and Unsolvability*, McGraw-Hill, New York.

- (1966a), “Diophantine Equations and Recursively Enumerable Sets” dans [**Caianiello E. R. ed.**, *Automata Theory*, Academic Press, London], pp. 146-152.

- (1966b), “Recursive Functions - an Introduction” dans [**Caianiello E. R. ed.**, *Automata Theory*, Academic Press, London], pp. 153-163.

Gandy R. O. (1969), "The Concept of Computability" dans [**Harré R. ed.** (1969), *Scientific Thought 1900-1960, a selective survey*, Clarendon Press, Oxford], pp. 1-4.

Kreisel G. (1974), "A Notion of Mechanistic theory", Synthese, vol. 29, pp. 11-26.

Kugel P.,

- (1986), "Thinking may be more than computing", Cognition, Vol. 22, n°2, pp. 137-198.
- (1987), "When is a computer not a computer ?", Cognition, Vol. 23, n°1, pp. 89-94.

4. Cybernétique, Intelligence artificielle

Andler D. (1990), "Quelle est la place de l'intelligence artificielle dans l'étude de la cognition ?", Revue internationale de philsosophie, Presses Universitaires de France, Paris, volume 44, n°172, 1/1990, pp. 62-86.

Dahlhaus E. et **Makowsky J. A.** (1988), "Gandy's Principles of Mechanisms as a Model of Parallel Computation" dans [**Herken R. ed.** (1988), *The Universal Turing machine*, Oxford University Press, 1988], pp. 309-314.

Delahaye J. -P. (1986), *Outils logiques pour l'intelligence artificielle*, Eyrolles, Paris.

Dupuy J.-P.,

- (1985), "L'essor de la première cybernétique" (1943-1953)", Cahiers du Crea, 7, Ecole Polytechnique, Paris.
- (1994), *Aux origines des sciences cognitives*, La Découverte, Paris.

Ladrière J. (1991), "L'intelligence artificielle", Les Etudes, tome 374, n°6 (3746), Juin, pp.777-788.

McCulloch W. S. et Pitts W. H. (1943), "A logical Calculus of Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics, volume 5, 1943 réédité dans [**Boden M. A. ed.** (1990), *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford], pp. 22-39.

Partridge D. et **Wilks Y. eds.** (1990), *The Foundations of Artificial Intelligence*; a sourcebook, Cambridge University Press, Cambridge.

Ruyer R. (1954), *La Cybernétique et l'Origine du Concept d'Information*, Flammarion, Paris.

Schank R. C. (1990), "What is AI anyway ?" dans [**Partridge D.** et **Wilks Y. eds.** (1990), *The Foundations of Artificial Intelligence*; a sourcebook, Cambridge University Press, Cambridge], pp. 3-13.

Schnelle H. (1988), "Turing naturalized : von Neumann's unfinished project" dans [**Herken R. ed.** (1988), *The Universal Turing machine*, Oxford University Press, pp. 539-559].

Winston P. H. (1977), *Artificial Intelligence*, Addison-Wesley.

5. <u>Linguistique</u>

Chambreuil et **Pariente J.-C.** (1990), *Langue naturelle et logique*, Peter Lang, Berne.

Cook V. J. (1988), *Chomsky's universal grammar*, Basil Blackwell, Oxford.

Hagège C. (1986), *L'homme de paroles*, Contribution linguistique aux sciences humaines, Gallimard, Paris.

Llorach E. (1968), "Les représentations graphiques du langage", dans [**Martinet A. dir.** (1968) *Le langage,* Encyclopédie de la Pléiade, Gallimard, Paris], pp. 513-568.

Martinet A. dir. (1968), *Le langage,* Encyclopédie de la Pléiade, Gallimard, Paris.

Petitot J. (1991), "Syntaxe topologique et grammaire cognitive", <u>Langages</u>, n° 103, pp. 97-128.

6. <u>Cognition</u>

Andler D.,

- (1986), "Le Cognitivisme orthodoxe en question", <u>Cahiers du Crea</u>, 9, Ecole

Polytechnique, Paris, pp. 9-106.

- (1987), “Logique, pensée, machine”, Esprit, n° 128, pp. 26-39.

- (1987), “Progrès en siuation d'incertitude”, Le Débat, n° 47.

- (1989), “Cognitives (sciences)” dans [Encyclopedia Universalis].

Boden M. A. ed. (1990),*The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford.

Bruter C. P. (1974- 1976-1986), *Topologie et perception*, trois vol., Maloine, Paris.

Cataneda H. N. ed. (1967), *Intentionality, Minds and Perception*, Wayne State University Press.

Chandrasekaran B. (1990)**,** “What kind of information processing is intelligence ?” dans [**Partridge D.** and **Wilks Y. eds.** (1990), *The Foundations of Artificial Intelligence; a Sourcebook*, Cambridge University Press, Cambridge], pp. 14-46.

Changeux J.-P. (1983), *L’homme neuronal*, Fayard, Paris.

Dreyfus H. L. (1991), “L’épiphénoménologie de Husserl”, traduction française dans Les Etudes philosophiques, n° 1, pp. 57-77.

Dreyfus S. et **Dreyfus H.** (1990), "Towards a reconciliation of phenomenology and AI", dans [**Partridge D.** and **Wilks Y. eds.** (1990),*The foundations of artificial intelligence; a sourcebook*, Cambridge University Press, Cambridge], pp. 396-410.

Fodor J.A. (1985), “Précis of *The Modularity of Mind*”, The Behavioral and Brain Sciences, 8, pp. 1-40.

Hofstadter D. (1987), “Cognition et subcognition, Sortir du rêve de Boole”, Le Débat, n° 47.

Jackendoff R. (1987), *Consciousness and the computational mind*, MIT Press, Cambridge, Massachussetts.

Jurdant B. (1984), “Ecriture alphabétique et stratégies cognitives”, Cahiers STS, n°5, “Querelle de modèles”, CNRS, Paris, pp. 92-102.

Langley P. et al. (1987), *Scientific discovery*, The MIT Press, Cambridge, Massachussetts.

McIntyre R. (1991), “Husserl et la théorie représentationnelle de l’esprit”, Topoi, 5, 1986; traduction française dans Les Etudes Philosophiques, n°1, pp. 31-56.

Mosconi J. (1991),"Sur quelques capacités et incapacités des machines", conférence faite à la Société Française de Philosophie, séance du 13 avril 1991, Bulletin de la société française de philosophie, Juillet-Septembre , 85 ème année, n°3, Armand Colin, Paris, pp. 82-115.

Newell A. et **Simon H. A.** (1976), "Computer Science as Empirical Enquiry : Symbols and Search", Communications of the Association for Computing Machinery, 19, réédité dans [**Boden M. A. ed.** (1990),*The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990, pp. 105-132].

Partridge D. et **Wilks Y. eds.** (1990), *The Foundations of Artificial Intelligence*; a sourcebook, Cambridge University Press, Cambridge.

Piatelli-Palmarini M. ed. (1979), *Théorie du langage, théorie de l'apprentissage*; le débat entre Jean Piaget et Noam Chomsky; Centre de Royaumont, Le Seuil, Paris.

Proust J. (1987), "L'intelligence artificielle comme philosophie", Le Débat, n°47.

Putnam H.,

- (1963), "Brains and Behavior" dans [**Putnam H.** (1975), *Philosophical Papers*, II, Cambridge University Press, Cambridge], pp. 325-341.
- (1967), "The mental life of some machines" dans [**Cataneda H. N. ed.** (1967), *Intentionality, Minds and Perception*, Wayne State University Press], pp. 84-100.

Rucker R. (1982), *Infinity and the Mind,* Birkhaüser.

Van Gelder T. (1992), "What might Cognition be if not Computation?", research report 75, Indiana University.

7. Psychologie, psychanalyse

Anzieu D.,

- (1985), *Le Moi-peau*, Dunod, Paris.
- (1993), *Les contenants de pensée*, Dunod, Paris.

Charraud N., (1994) *Infini et inconscient; Essai sur Georg Cantor*, Anthropos-

Economica, Paris.
Gadet F., **Haroche C.**, **Henry P.** et **Pêcheux M.** (1982), “Notes sur la question du langage et du symbolique en psychologie”, Fundamenta Scientiae, vol 3, n°2, pp. 149-159.
Guillaume P. (1937), *La psychologie de la forme*, Flammarion, Paris.
Köhler W. (1971), *The selected papers of Wolfgang Köhler*, Henle M. ed., Liveright Publishing Co.

8. Epistémologie générale

Arsac J. (1987), *Les machines à penser*, Le Seuil, Paris.
Atlan H. (1979), *Entre le cristal et la fumée : essai sur l’organisation du vivant*, Le Seuil, Paris.
Baccou R. (1951), *Histoire de la science grecque de Thalès à Socrate*, Aubier-Montaigne, Paris.
Badiou A. (1988), *L'être et l'événement*, Le Seuil, Paris.
Bouligand Y. ed. (1980), *La Morphogenèse*, Maloine, Paris.
D'Arcy Thompson (1917), *On Growth and Form*, Cambridge University Press, Abridged ed. 1961.
Dawkins R. (1986), *The Blind Watchmaker*, Norton, New York and London.
Delattre P. (1971), *Système, structure, fonction, évolution*, Maloine, Paris.
Gödel K. (1949), “A Remark about the Relationship between Relativity Theory and Idealistic Philosophy”, dans [**Schilpp P. A. ed.** (1949), *Albert Einstein : Philosopher-Scientist*, vol. 2, The Library of Living Philosophers, Open Court, La Salle, Illinois], pp. 555-562.
Granger G.-G.,
- (1984), “Modèles qualitatifs, modèles quantitatifs dans la connaissance scientifique”, Cahiers STS n°5, “Querelle de modèles”, CNRS, Paris, pp. 11-18.
- (1988), *Pour la connaissance philosophique* , ed. Odile Jacob, Paris.

Hao Wang (1988), *Beyond analytic philosophy; doing justice to what we know*, MIT Press, Cambridge, Massachussetts.

Harré R. ed. (1969), *Scientific Thought 1900-1960, a selective survey*, Clarendon Press, Oxford.

Largeault J.,

- (1971), *Enquête sur le nominalisme,* Editions Nauwelaerts, Louvain.
- (1984), *Leçons de métaphysique,* Ed. Université de Paris-Val de Marne.
- (1985), *Systèmes de la nature,* Vrin, Paris.
- (1988), *Principes classiques d'interprétation de la nature*, Publications de l'Institut Interdisciplinaire d'Etudes Epistémologiques, Vrin, Paris.

Levine R. et **Tribus M. eds.** (1979),*The Maximum Entropy Formalism*, MIT Press, Cambridge, Massachussetts.

Meigne M.,

- (1963), *Structure de la Matière*, Presses Universitaires de France, Paris.
- (1964), *Recherches sur une logique de la pensée créatrice en mathématiques*, Blanchard, Paris.

Petitot J.,

- (1987), "Schématisme et interprétation ou La logique transcendantale comme ontologie herméneutique", <u>Confrontation</u>, n°17, Aubier, pp. 41-71.
- (1988), "Structuralisme et Phénoménologie : la théorie des catastrophes et la part maudite de la raison" dans [**Petitot J. dir.**, *Logos et théorie des catastrophes*, à partir de l'œuvre de René Thom, Actes du Colloque de Cerisy de 1982, Patiño, Genève], pp. 345-376.
- (1988), **Petitot J. dir.**, *Logos et théorie des catastrophes*, à partir de l'œuvre de René Thom, Actes du Colloque de Cerisy de 1982, Patiño, Genève.

Philonenko A. (1990), *Le transcendantal et la pensée moderne*, Paris, Presses Universitaires de France.

Poincaré H., (1913), *Dernières pensées*, Flammarion, Paris.

Putnam H.,

- (1981), *Reason, Truth and History,* Cambridge University Press; traduction française *Raison, Vérité et Histoire*, Editions de Minuit, Paris, 1984.
- (1988), *Representation and reality*, MIT Press; traduction française

Représentation et réalité, traduction française, Gallimard, Paris, 1990.

Salanskis J.-M.,

- (1990), "Le concept de Gestalt et la situation contemporaine de la philosophie des sciences", Les Etudes philosophiques, n°4, 1990, pp.519-536.

- (1992), "Herméneutique, Logique, Mathématique", Les Etudes philosophiques, n°2, pp. 257-273.

- (1993), "Modèles du continu dans les sciences", Intellectica, 1/2, 13-14, pp. 45-78.

Schilpp P. A. ed. (1949), *Albert Einstein : Philosopher-Scientist*, vol. 2, The Library of Living Philosophers, Open Court, La Salle, Illinois.

Scubla L. (1993), "Classification des sciences et philosophie de la nature; Prolégomènes à une épistémologie des sciences de l'homme et de la société", Cahiers du Crea, n°15, Ecole Polytechnique, Paris, pp. 49-91.

Thom R.,

- (1988), *Esquisse d'une sémiophysique,* InterEditions, Paris.

- (1990), *Apologie du logos*, Hachette, Paris.

Vuillemin J.,

- (1955), *Physique et métaphysique kantiennes*, Presses Universitaires de France, Paris.

- (1960), *Mathématiques et métaphysique chez Descartes,* Presses Universitaires de France, Paris.

- (1985), "Sur deux cas d'application de l'axiomatique à la philosophie : l'analyse du mouvement par Zénon d'Elée et l'analyse de la liberté par Diodore Kronos", Fundamenta Scientiae, vol 6, n° 3, pp. 209-219.

Table des matières